Extension of the Hyper Run Assignment Model to real-time passengers forecasting in congested transit networks based on count data.

Lory Michelle Bresciani Miristice*[1], Guido Gentile[1], Daniele Tiddi[2], and Lorenzo Meschini[2]

[1]DICEA, University of Rome La Sapienza, Via Eudossiana 18, Rome, Italy

[2]PTV Group SISTeMA, Via Spallanzani 14, Rome, Italy

SHORT SUMMARY

Recurrent and non-recurrent congestion phenomena increasingly affect densely interconnected transit networks. In particular, the measures adopted to contain the spread of the COVID-19 pandemic significantly affect public transport capacity, increasing congestion. Typical congestion phenomena, together with service disruptions and atypical demand, can lead to low levels of service harming planned schedules. Therefore, transit operators require a tool that can quickly forecast a potential lack of capacity in transit systems, to perform service recovery (e.g., introducing new runs) and inform passengers about crowding (e.g., through real-time information panels or trip planners). This research proposes an innovative congested run-based macroscopic dynamic assignment model that incorporates real-time measurements and events to compute users' elastic route choices under the assumption that passengers are fully informed. The model simulates the effects of congestion events and countermeasures introduced by the operators, allowing them to test several scenarios on large transit networks faster than in real-time.

Keywords: implicit hyperpaths, public transport services, real-time data, schedule-based assignment, short-term forecast, vehicle capacity constraints.

1. INTRODUCTION

Recurrent congestion phenomena and unexpected events (e.g., abnormal demand fluctuations, complete line failures) affect passengers' route choices and the propagation of their flow, leading to service degradation. Public transport agencies need a model for optimal real-time transit management to mitigate the negative impact on the rest of the network. This model must be able to quickly predict the distribution of passengers across the network and provide volumes of passengers on specific runs, taking into account: 1) relevant congestion phenomena (i.e., overcrowding and strict capacity constraints); 2) real-time measurements (i.e., passenger counts and vehicle locations); and 3) real-time events (e.g., stop closures and run cancellations). Several researchers have addressed the issue of short-term ridership forecasting in recent decades, mainly focusing on data-driven demand forecasting (e.g., Zhang et al. [2011], Ma et al. [2014], Xue et al. [2015], Ding et al. [2016], Zhang et al. [2017], Zhang et al. [2020]). Liu and Chow [2021] proposes a dynamic passenger flow estimator to predict origin-destination demand and line flows using station count data in a congested schedule-based User Equilibrium (UE) model. However, this estimator is unsuitable for optimal real-time transit management because it does not account for real-time service disruptions.

We propose a model that predicts short-term passenger flows based on real-time schedules and passenger counts by performing an online Dynamic Transit Assignment (DTA). The model extends the Hyper Run Assignment Model (HRAM) proposed by Gentile et al. [2021] and considers real-time measurements and events in a Rolling-Horizon (RH) framework. The model: 1) computes elastic route choices by building the diachronic graph of HRAM starting from the Estimated Time of Arrival (ETA) of vehicles at stops; and 2) corrects the volumes resulting from flow propagation using passenger counts (as Meschini and Gentile [2009], Gentile and Meschini [2011], Attanasi et al. [2015] and Kucharski and Gentile [2019] do for road networks).

The contribution of this work is the use of HRAM to perform online DTA for short-term ridership forecasting, continuously updating the underlying service schedule and demand flows to account for up-to-date data coming from the field.

The social distancing measures affecting public transport usually adopted to limit the spread of the COVID-19 pandemic imply a capacity reduction of transit vehicles on the supply side and the introduction of staggered working hours to spread peaks on the demand side. The features of the proposed model should then become fundamental for transit operators, who would quickly forecast real-time overcrowding issues and adopt mitigating solutions (e.g., introduce new runs, inform passengers).

2. METHODOLOGY

HRAM [Gentile et al., 2021] simulates the decision of how to continue a trip to the destination when a fail-to-board event occurs (i.e., when passengers cannot board a vehicle because there is not enough space left). Since the outcome of these events is unknown in advance, a passenger chooses a strategy (rather than a simple path) before starting the trip, aiming to minimize the expected route cost (which depends on the flow pattern in the framework of dynamic user equilibrium). HRAM uses hyperpaths (introduced by Nguyen and Pallottino [1988]) in a schedule-based framework to represent strategies connected to fail-to-board probabilities (firstly proposed by Hamdouch and Lawphongpanich [2008]). It performs DTA of a given transit supply and dynamic demand by solving the fixed-point problem, adopting a sequential route choice model (proposed by Gentile and Papola [2006]) and a Gradient Projection (GP) algorithm for the solution of the UE (which improves the convergence on highly congested networks as shown by Gentile [2016]). This approach does not require formalizing entire trips and allows an implicit representation of hyperpaths, enhancing run-times that are compatible with real-time applications in RH.

HRAM describes the transit supply as a space-time directed graph (Figure 1), where each arc represents a different phase of the passengers' trip and connects two nodes with increasing time. With this approach, the diachronic graph is inherently acyclic, and the computation of the shortest tree requires a simple visit in reverse chronological order of all nodes (computed in linear time, as shown by Gentile [2017]). Moreover, the model directly provides volumes of passengers on specific runs (which is a requirement of real-time transit management).

HRAM represents strict capacity constraints by introducing a fail-to-board hyperarc for each departure N_{rs}^{depart} of run $r \in R_l$ (the set of runs of line $l \in L$) at stop $s \in S$. Each fail-to-board hyperarc is bifurcated and consists of a diversion node N_{rs}^{board}, a fail arc $(N_{rs}^{board}, N_{rs}^{alight})$, and a board arc $(N_{rs}^{board}, N_{rs}^{depart})$. At each event-based diversion node encountered by the passenger, the strategy associates a branch of the corresponding hyperarc to each possible outcome. Note that the conditional probability of each hyperarc branch is not the result of a choice but the outcome of the fail-to-board event (passengers take routing choices mainly at stop nodes). Fail-to-board probabilities π_a (i.e., the probability that a passenger waiting for a run on a crowded platform fails to get on the arriving vehicle) are computed as follows, assuming that all mingling passengers have equal probability to board (unlike in Trozzi et al. [2013] and Trozzi et al. [2015] where there is

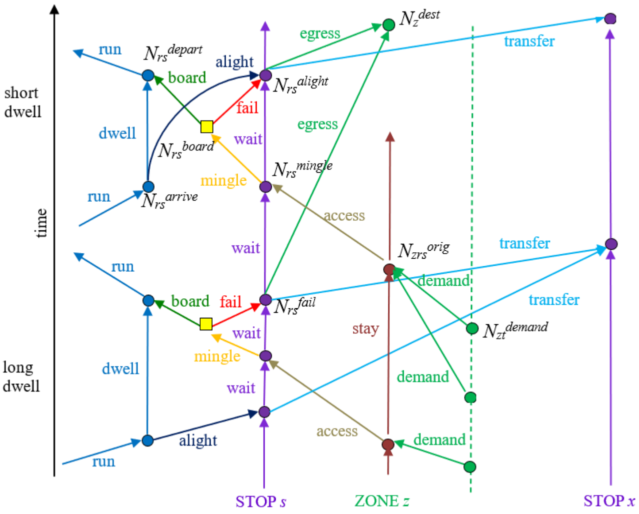


Figure 1: HRAM diachronic graph schema Gentile et al. [2021]. Transit service is organized in a set of line L, of stops S and runs R. A run $r \in R_l$ is a vehicle serving the stop sequence S_l of a line $l \in L$ according to a given timetable and is described by the following arcs: alight, dwell, mingle (i.e., passengers trying to board), fail, board and run. Stop $s \in S_l \subseteq S$ is the only place where passengers can board and alight, and we assume that a positive dwelling time $t_r s^{dwell}$ is needed for boarding and alighting operations. N_{rs}^{board} (yellow square) represents the diversion node of the fail-to-board hyperarc, introduced to model strict capacity constraints. The centroid of each zone $z \in Z$ is represented by multiple origin nodes, one for each possible departure from a connected stop, and one destination node. Demand nodes and arcs are introduced to model the possibility of delaying the departure with respect to the desired time, using the stay arcs at origins. Finally, access, egress and transfer arcs represent direct connections between stops and zones, and among stops themselves. They are added only for significant events (run arriving/departing from a stop) and their topology (and thus the travel time) depends on the walking shortest time between the connected elements.

queue with priority). Being the probability of boarding the run:

$$\pi_{ag} = \min\left(1, \frac{k_r - q_b}{q_a}\right) \quad \forall a \in A_r^{board}, b = A_a^{dwell}, \tag{1}$$

the fail-to-board probability is the ones' complement of π_{ag}:

$$\pi_{bg} = 1 - \pi_{ag} \quad \forall b \in A_r^{fail}, a = A_b^{board}, \tag{2}$$

where:

- A_a^{dwell} is the dwell arc associated with the board arc $a \in A_r^{board}$;
- A_b^{board} is the board arc associated with the fail arc $b \in A_r^{fail}$;
- k_r is the run capacity.

We have extended HRAM to consider real-time data and use it to perform online DTA, as shown in Figure 2. The model computes online DTA for dynamic demand and supply (the space-time network of Figure 1 built from ETA at stops). It accounts for real-time passenger counts and events by 1) using HRAM to compute elastic local users' decisions p_a (arc conditional probabilities); 2) running a Flow Propagation Model (FPM) to load the demand consistently with the arc conditional probabilities p_a computed by HRAM. FPM considers strict capacity constraints by using fail-to-board probabilities π_a, limiting the volume propagated to board arcs and forcing the remaining passengers to the corresponding fail arcs. FPM uses observed volumes (passenger counts) to correct errors in flow propagation, by overwriting the propagated volumes with the observed ones.

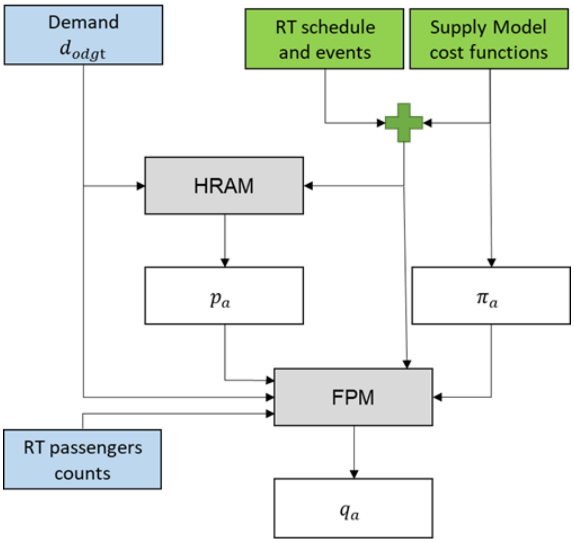


Figure 2: Online DTA schema. The model performs online DTA for dynamic demand d_{odgt} and dynamic supply (the space-time network representing schedule-based services and their capacities updated with ETA at stops), taking into account real-time passenger counts and events. It: 1) uses HRAM to compute local users' decisions p_a (arc conditional probabilities); 2) runs a Flow Propagation Model (FPM) with strict capacity constraints (using fail-to-board probabilities π_a) that loads demand consistently with the arc conditional probabilities computed by HRAM. Passenger counts (observed volumes) are used in FPM to correct the propagated volumes by overwriting them in the corresponding arcs.

We run online DTA in an RH framework to account for up-to-date data (as Gentile et al. [2013] does for dynamic traffic assignment), solving the deterministic problem iteratively by performing sequential online DTAs (one for each iteration). Each online DTA simulates a prediction horizon, assuming that passengers know all events related to this time horizon no later than their start time.

Each iteration starts with the volumes resulting from the previous simulation ("warm start"), recovered using a Flow Recovery Model (FRM). First, HRAM computes the elastic local user decisions on the real-time supply initializing: 1) its route choice model with the expected cost of reaching the destinations from each stop at the end of the current prediction horizon (recovered from an offline DTA with a longer time horizon); 2) its flow propagation model with the demand flow for each node at the beginning of the current prediction horizon (recovered from the volumes resulting from the previous simulation). Then, the Flow Propagation Model (FPM) uses the elastic local user decisions computed by HRAM to forward the demand and the aggregated flows of the previous state in chronological order, starting from the origin nodes.

Flow recovery model

Both HRAM and FPM require restoring the previous simulation state (i.e., diachronic graph and volumes q_a^* on each arc). Specifically, HRAM requires the recovery of node destination flows q_i^s from the $|Z|$ simulation states resulting from the previous simulation HRAM, while FPM requires the recovery of aggregated node flows q_i^s and aggregated arc flows q_a^s from the simulation state resulting from the previous simulation FPM. FRM is the same for both HRAM and FPM and accounts for schedule changes due to real-time variations in transit service.

FRM recovers the flows q_a^* from the previous simulation state for all arcs $a \in A$ that started before the start time $\tilde{\tau}$ of the current simulation. Then, it initializes the flows required by HRAM and FPM as follows:

$$q_a^s = q_a^* \quad \forall a \in A \mid \tau_{N_a^-} < \tilde{\tau}, \tag{3}$$

$$q_i^s = \textstyle\sum_{a \in A_i^+} q_a^* \quad \forall i \in N, \tag{4}$$

where:

- q_a^s and q_i^s are the flows restored for arc $a \in A$ and node $i \in N$;
- N_a^- is the tail node of arc $a \in A$ and $\tau_{N_a^-}$ is its clock-time;
- A_i^+ is set of arcs exiting node $i \in N$.

For arcs of type wait, a flow δ_{ab} (i.e., volume variation due to delays/advances of ETA at stops) is added/subtracted to the resulting flow q_a^* to account for real-time schedule changes.

Flow propagation Model

Flow q_i of node $i \in N$ is propagated forward in chronological order starting from origin nodes, as follows:

$$q_i = d_i + \textstyle\sum_{a \in A_i^-} q_a \quad \forall i \in N, \tag{5}$$

$$q_a = q_a^m \quad \forall a \in A^m, \tag{6}$$

$$q_a = q_a^s \quad \forall a \in A^s \setminus A^m, \tag{7}$$

$$q_a = q_i \cdot p_a \quad \forall a \in A_i^+, a \notin A^s \cup A^m, \tag{8}$$

where:

- A_i^-/A_i^+ is the set of arcs entering/exiting node $i \in N$;
- $A^m \subseteq A$ is set of arcs with measurements (i.e., observed volumes);
- $A^s \subseteq A$ set of arcs with flow restored from the simulation previous state;
- q_a is the flow on arc $a \in A$;
- d_i is the demand (number of passengers) departing from node $i \in N$ (null for all nodes except for each demand node $i = N_{zt}^{demand}$);
- q_a^m is the measured flow on arc $a \in A^m$ (obtained by converting the real-time passenger counts to passengers volume on the associated arcs);
- q_a^s is the flow on arc $a \in A^s$ obtained from the state snapshot;

- p_a is the aggregated arc turn probability for arc $a \in A$.

Aggregated arc turn probabilities p_a are calculated according to the arc type and the presence of measurements/snapshot flows, as follows:

$$p_a = 0 \quad \forall a \in A^s \cup A^m, \tag{9}$$

$$p_a = \pi_a^{board} \quad \forall a \in A^{board}, \tag{10}$$

$$p_a = \pi_a^{fail} \quad \forall a \in A^{fail}, \tag{11}$$

$$p_a = p_a^{HRAM} \quad \forall a \notin A^s \cup A^m,\ a \notin A^{board},\ a \notin A^{fail}, \tag{12}$$

where:

- π_a^{board} is the boarding probability of arc $a \in A^{board}$ computed by (1);
- π_a^{fail} is the failing-to-board probability of arc $a \in A^{fail}$ computed by (2);
- p_a^{HRAM} is the aggregated arc conditional probability of arc $a \in A$ resulted from HRAM.

Elastic turn probabilities may be null for all the arcs exiting a certain node (passenger counts on arcs not loaded by HRAM, or unexpected failing-to-board in the FPM). In this case, they are calculated proportionally to the number of arcs in the forward star of the node.

3. RESULTS AND DISCUSSION

This section describes the numerical tests conducted, which are: 1) the validation against service variations on a toy network; 2) the evaluation of algorithm performances and inclusion of passengers' count on a medium-size real network.

Validation against service variations

We have tested the proposed model against service variations on the toy network in Figure 3, using the transit service of Table 1 and the travel demand of Table 2 (passengers leave their origin at 8:00).

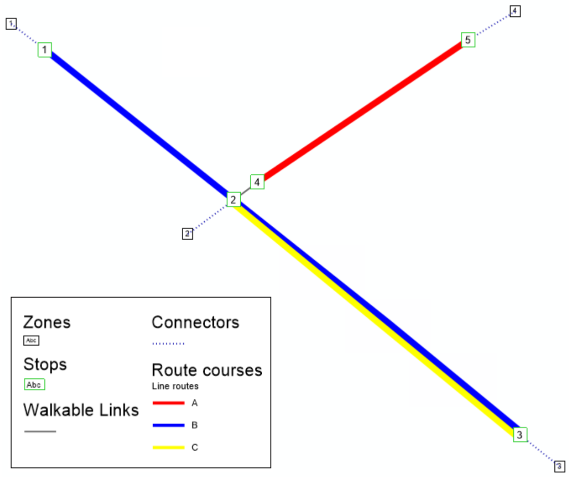


Figure 3: Toy Network. The network consists of four zones connected by three line routes. *Line route B* serves *Zone 1*, *Zone 2* and *Zone 3*. *Line route C* serves *Zone 2* and *Zone 3*. *Line route A* serves *Zone 4*. Passengers can walk from the stop related to *Zone 2* (i.e., *Stop 2*) to the initial stop of *Line route A* (i.e., *Stop 4*).

Table 1: Toy network, timetable of transit service.

Run	Vehicle Capacity	Arrival Times				
		Stop 1	*Stop 2*	*Stop 3*	*Stop 4*	*Stop 5*
B1	150	08:00	08:06	08:12	-	-
C1	50	-	08:06	08:12	-	-
B2	150	08:10	08:16	08:22	-	-
A1	50	-	-	-	08:20	08:25
B3	150	08:30	08:36	08:42	-	-
C2	50	-	08:30	08:36	-	-
A2	50	-	-	-	08:30	08:35

Table 2: Toy network, origin-destination matrix.

Demand	*Zone 1*	*Zone 2*	*Zone 3*	*Zone 4*
Zone 1	/	50	50	50
Zone 2	/	/	300	/
Zone 3	/	/	/	/
Zone 4	/	/	/	/

Starting from 7:30, we have run simulations of one hour every 30 minutes. Figure 4 shows the flow distribution obtained from the first simulation (*Simulation 1*, 7:30 - 8:30). In particular, 150 passengers were unable to board their desired run (i.e., *B1* and *C1*) at *Stop 2* and waited for the next available run (i.e., *B2*). At 8:00 (i.e., the start time of the second simulation), 150 passengers were boarding run *B1* at *Stop 1*, while the remaining 300 passengers were waiting at *Stop 2* for the first available run.

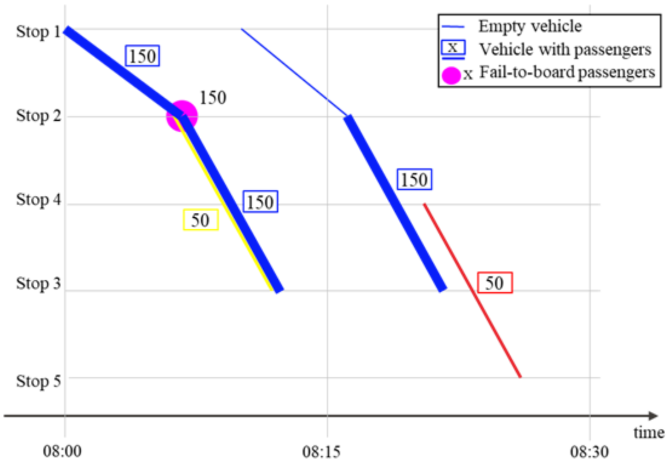


Figure 4: Toy Network, resulting flow distribution of *Simulation 1*. The first simulation (7:30 - 8:30) showed that 150 passengers were unable to board their desired run (i.e., *B1* and *C1*) at *Stop 2* and waited for the next available run (i.e., *B2*). At 8:00 (i.e., the start time of the second simulation), 150 passengers were boarding run *B1* at *Stop 1*, while the remaining 300 passengers were waiting at *Stop 2* for the first available run.

We have run the second simulation (*Simulation 2*, 8:00 - 9:00) in the different scenarios of Table 3.

Table 3: Toy network, scenarios tested for *Simulation 2* (8:00-9:00).

Scenario	**Description**
1	no measurements/events
2	run *C1* delayed 4 minutes
3	run *C1* delayed 10 minutes
4	run *C1* and *C2* cancelled
5	run *A1* cancelled
6	stop *2* of run *B1* disabled

Figure 5 shows the flow distribution obtained from these scenarios. The results were consistent with the expected behavior: passengers reroute when possible. Scenario 6 (Figure 5f) shows an example of impossible rerouting. Some passengers boarded run *B1* at *Stop* 1 in the first simulation, planning to get off at *Stop 2*. In the second simulation, those passengers found out that *Stop 2* was closed and had to stay over the vehicle, even though it did not bring them to their destination.

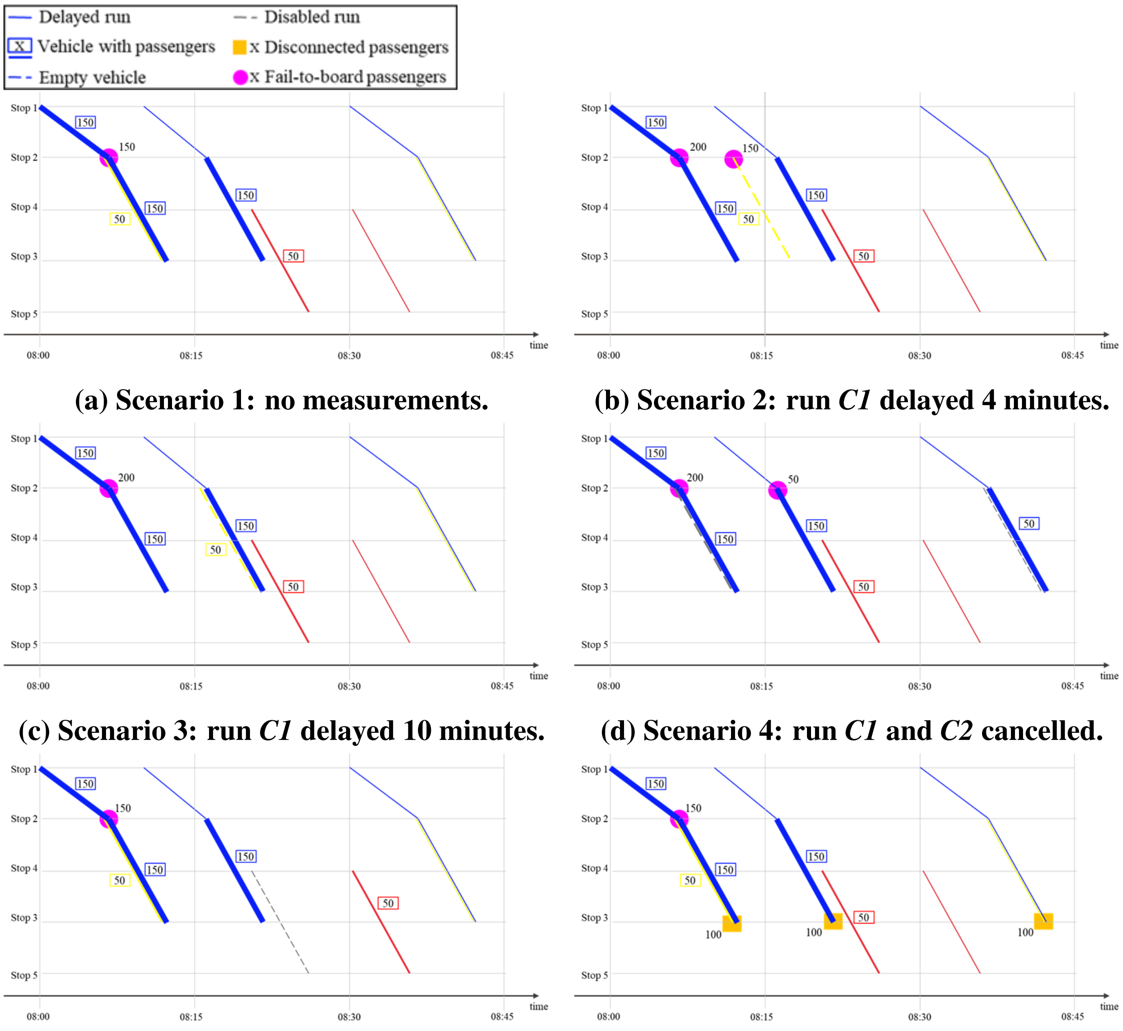
Performance and passengers' count inclusion on medium-sized real network

We have tested the proposed model on a medium-sized network, having: 123,130 streets, 6,748 stops, 1,417 line routes, and 119,661 runs throughout the day. We have run simulations every 5 minutes with a prediction horizon of 1 hour, simulating the morning peak hour (6:00 to 9:00). We have used an ordinary workstation machine with an Intel(R) Xeon(R) Platinum processor and 16GB of RAM to run the test. Each simulation performed five iterations of HRAM to compute elastic route choices and took about 83 seconds, which is in line with computation times required by real-time management.

We have tested the inclusion of observed volumes (i.e., passenger counts) in FPM by adding a measurement of 5000 passengers on a given subway platform at 07:05 to simulate an atypical demand due to some event (e.g., a sporting event). In particular, this platform had 10 passengers waiting for the subway between 07:05 and 07:10 (results are with a discretization of 5 minutes). Once adding the measurement, we have confirmed that simulations starting after 07:05 had 5000 passengers waiting at the platform from 07:05 to 07:10. Considering that the subway capacity is 1,200 passengers, not all passengers could board the first arriving vehicle. We have also observed congestion spreading all over the network, starting from the platform with extra demand and propagating in space and time.

4. CONCLUSIONS

Recurrent and non-recurrent congestion phenomena increasingly affect densely interconnected transit networks. In particular, the measures adopted to contain the spread of the COVID-19 pandemic significantly affect public transport capacity, increasing congestion. Typical congestion phenomena, together with service disruptions and atypical demand, can lead to low levels of service harming planned schedules. Therefore, transit operators require a tool that can quickly forecast a potential lack of capacity in transit systems, to perform service recovery (e.g., introducing new runs) and inform passengers about crowding (e.g., through real-time information panels or trip planners). This research presents a model that extends the Hyper Run Assignment Model



(a) Scenario 1: no measurements.

(b) Scenario 2: run *C1* delayed 4 minutes.

(c) Scenario 3: run *C1* delayed 10 minutes.

(d) Scenario 4: run *C1* and *C2* cancelled.

(e) Scenario 5: run *A1* cancelled.

(f) Scenario 6: stop *2* of run *B1* disabled.

Figure 5: Toy network, flow distribution obtained from tested scenarios on events. In scenario 1 there are no variations in the timetable and passengers confirm the choice of *Simulation 1*. In scenario 2 and 3, run *C1* was delayed but passengers were able to confirm their choices (even if with some delays). In scenario 4, run *C1* and *C2* were cancelled and passengers rerouted to run *B2* and *B3*. In scenario 5, run *A1* was cancelled and passengers rerouted to run *A2*. In scenario 6, stop *2* of run *B1* was disabled and 100 passengers were unable to reroute. Those passengers boarded run *B1* at *Stop* 1 at 8:00 (*Simulation 1*) planning to get off at *Stop 2*, but they found out that *Stop 2* was closed (*Simulation 2*) and had to stay over the vehicle, even though it did not bring them to their destination.

(HRAM) proposed by Gentile et al. [2021] to predict short-term passenger flows based on real-time schedules and passenger counts.

The proposed model considers real-time measurements and events by performing sequential online DTAs that overlap partially (RH). Each simulation starts from a "warm start" considering the flows resulting from the previous DTA. Every online DTA: 1) extends HRAM to compute elastic route choices taking into account real-time data (i.e., ETA at stops and events), assuming fully informed passengers; and 2) propagates the demand according to these choices, overriding the propagated flows when real-time passenger counts are available. The model adopts fail-to-board hyperarcs to simulate strict capacity constraints and provides volumes of passengers on specific runs (which is a requirement for real-time operation). HRAM adopts a GP algorithm to reach convergence, as strict capacity constraints hinder the convergence of simpler algorithms (e.g., method of successive averages). HRAM has running times compatible with real-time management thanks to implicit hyperarc enumeration.

We have validated the model against service variations on a toy network. We have shown that passengers can reroute if they discover the event (i.e., service variation) before making their final choice (i.e., passengers cannot change the desired run if they are already on board the one affected by the event). Moreover, we have tested the algorithm performances on a real network, obtaining computation times in line with the ones required by real-time management. We have also verified the inclusion of passenger counts on the same network, adding measurement and observing the related congestion spreading all over the transit service.

At its current state, the proposed model computes elastic route choices through HRAM, assuming that passengers are aware of all events related to the prediction horizon no later than their start time. However, the model could include different levels of information to relax the assumption of fully informed passengers. Furthermore, the model does not account for passenger counts when computing elastic route choices. Including them is straightforward and can be achieved using the observed volumes when updating the costs observed by the passengers. Forthcoming papers will address the proposed improvements.

REFERENCES

A. Attanasi, E. Silvestri, P. Meschini, and G. Gentile. Real world applications using parallel computing techniques in dynamic traffic assignment and shortest path search. In *IEEE 18th International Conference on Intelligent Transportation Systems*, pages 316–321, Gran Canaria, Spain, 09 2015.

C. Ding, D. Wang, X. Ma, and H. Li. Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8(11), 2016.

G. Gentile. Solving a dynamic user equilibrium model based on splitting rates with gradient projection algorithms. *Transportation Research Part B: Methodological*, 92:120–147, 10 2016.

G. Gentile. Time-dependent shortest hyperpaths for dynamic routing on transit networks. In A. Nuzzolo and W. Lam, editors, *Modelling Intelligent Multi-Modal Transit Systems*, pages 174–230. CRC Press Tailor & Francis, 11 2017.

G. Gentile and L. Meschini. Using dynamic assignment models for real-time traffic forecast on large urban networks. In *Proceedings of the 2nd International Conference on Models and Technologies for Intelligent Transportation Systems*, Leuven, Belgium, 06 2011.

G. Gentile and A. Papola. An alternative approach to route choice simulation: the sequential models. In *Proceedings of the European Transport Conference (ETC) 2006*, Strasbourg, France, 09 2006.

G. Gentile, R. Kucharski, and L. Meschini. Modelling rerouting phenomena through dynamic traffic assignment in rolling horizon. In *Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transportation Systems*, pages 513–522, 2013.

G. Gentile, L. M. B. Miristice, D. Tiddi, and L. Meschini. The hyper run assignment model: simulation on a diachronic graph of congested transit networks with fail-to-board probabilities at stops. In *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pages 1–7, 2021.

Y. Hamdouch and S. Lawphongpanich. Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B: Methodological*, 42:663–684, 08 2008.

R. Kucharski and G. Gentile. Simulation of rerouting phenomena in dynamic traffic assignment with the information comply model. *Transportation Research Part B: Methodological*, 126: 414–441, 2019.

Q. Liu and J. Y. J. Chow. A congested schedule-based dynamic transit passenger flow estimator using stop count data. *ArXiv*, abs/2107.08217, 2021.

Z. Ma, J. Xing, M. Mesbah, and L. Ferreira. Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C: Emerging Technologies*, 39:148–163, 2014.

L. Meschini and G. Gentile. Real-time traffic monitoring and forecast through optima - optimal path travel information for mobility actions. In *Proceedings of International Conference on Models and Technologies for Intelligent Transportation Systems*, Rome, Italy, 06 2009.

S. Nguyen and S. Pallottino. Equilibrium traffic assignment for large scale transit networks. *European Journal of Operational Research*, 37:176–186, 11 1988.

V. Trozzi, I. Kaparias, M. Bell, and G. Gentile. A dynamic route choice model for public transport networks with boarding queues. *Transportation Planning and Technology*, 36:44–61, 02 2013.

V. Trozzi, G. Gentile, I. Kaparias, and M. G. H. Bell. Effects of countdown displays in public transport route choice under severe overcrowding. *Networks and Spatial Economics*, 15(3): 823–842, 2015.

R. Xue, D. J. Sun, and S. Chen. Short-term bus passenger demand prediction based on time series model and interactive multiple model approach. *Discrete Dynamics in Nature and Society*, 2015, 2015.

C. H. Zhang, S. Rui, and Y. Sun. Kalman filter-based short-term passager flow forecasting on bus stop[j]. *Journal of Transportation Systems Engineering and Information Technology*, 11: 155–156, 08 2011.

J. Zhang, D. Shen, L. Tu, F. Zhang, C.-Z. Xu, Y. Wang, L. Ruyue, X. Li, B. Huang, and Z. Li. A real-time passenger flow estimation and prediction method for urban bus transit systems. *IEEE Transactions on Intelligent Transportation Systems*, PP:1–11, 04 2017.

J. Zhang, F. Chen, Z. Cui, Y. Guo, and Y. Zhu. Deep learning architecture for short-term passenger flow forecasting in urban rail transit. *IEEE Transactions on Intelligent Transportation Systems*, page 1–11, 2020.