Peak-hour Rail Demand Shifting with Discrete Optimisation

John M. Betts, David. L. Dowe, Daniel Guimarans, Daniel D. Harabor, Heshan Kumarage, Peter J. Stuckey, and Michael Wybrow

Faculty of Information Technology, Monash University, Melbourne, Australia
{firstname.lastname}@monash.edu

Abstract. In this work we consider an information-based system to reduce metropolitan rail congestion in Melbourne, Australia. Existing approaches aim to reduce congestion by asking commuters to travel outside of peak times. We propose an alternative approach where congestion is reduced by enabling commuters to make an informed trade-off between travel time and ride comfort. Our approach exploits the differences in train frequency and stopping patterns between stations that results in trains, arriving within a short time of each other, to have markedly different levels of congestion, even during peak travel periods. We show that, in such cases, commuters can adjust their departure and arrival time by a small amount (typically under 10 minutes) in exchange for more comfortable travel. We show the potential benefit of making this trade-off with a discrete optimisation model which attempts to redistribute passenger demand across neighbouring services to improve passenger ride comfort overall. Computational results show that even at low to moderate levels of passenger take-up, our method of demand shifting has the potential to significantly reduce congestion across the rail corridor studied, with implications for the metropolitan network more generally.

1 Introduction

Home to more than 4.8 million residents, Melbourne is Australia's second-largest, and fastest growing city. Melbourne residents enjoy access to an extensive public transportation network which includes metropolitan rail, light rail and bus services. According to Public Transport Victoria (PTV)[1] there were 565 million trips on public transport in the Melbourne metropolitan area in the year from 1 July 2017 – 30 June 2018 [2]. Of these, the largest share belongs to rail, with 240.9 million trips recorded. One of the major challenges facing transport planners in Melbourne is that rail passengers often experience high levels of congestion, especially during morning and afternoon peak periods. Attempts to tackle Melbourne's congestion tend to focus on the addition of more infrastructure, that is, rail lines and trains. However, this approach is expensive, and

[1] PTV is the government agency responsible for providing and coordinating public transport for Melbourne and across the state of Victoria.

sometimes impossible due to limitations on available space and other resources. Two alternative strategies, both widely studied in the research literature, are: (i) optimisation-based demand management and; (ii) demand management via incentives. We briefly discuss each.

Optimisation-based demand management, sometimes called passenger flow control, works by directing passengers to services based on their planned journey; e.g. [14, 8]. This body of research shows that passenger load can be effectively moved downstream until the demand in a saturated system is resolved. Disadvantages of this type of approach include high planning overheads, as passenger flows need to be optimised in near real time, and a dependence on significant physical infrastructure, such as waiting areas, boarding areas and designated entries. This approach also presumes that passengers will tolerate it.

Incentive-based demand management, by comparison, exploits trade-offs that exist between passenger preferences for time, comfort and cost. The idea is to encourage passengers to travel during periods of reduced demand and to discourage travel during periods of peak-demand. Studies in this area often apply equilibrium modelling, seeking to quantify, under certain conditions, the dis-utility of travelling early or late against the cost of discomfort and the willingness to pay [12, 17]. The main disadvantage of such in-principle economic models is that proposed fare structures are complicated and their actual effects on real schedules are usually not clear. When applied in practice, incentive-based systems employ more simplified structures. One example is PTV's *Early Bird train travel* [9] a scheme that allows Melbourne passengers to ride for free provided they arrive at their destination before 7:15am on weekdays. Another example is Singapore's INSINC [13] system, which rewards passengers who shift their travel away from periods of peak demand. These approaches report varying degrees of success but related studies [16, 9] show that relatively high reductions in fare are sometimes necessary to overcome the reluctance of some passengers to avoid peak periods.

In this work we consider a different approach where we aim to shift demand *within* peak periods by encouraging commuters to make informed trade-offs. We are motivated by evidence from the literature which suggests that passengers are willing to incur some additional travel time in order to secure a more comfortable trip [7, 6, 3, 5]. In the case of Melbourne, electronic noticeboards at rail stations, and also travel apps, show only the time of the next departure. However, many stations are serviced by multiple lines, including some with low levels of occupancy, even during peak periods. We posit that, if congestion information was made available, passengers could make informed trade-offs based on preferences and needs. For example, pregnant, elderly, or disabled passengers might prefer a longer seated trip to a shorter one that involves standing.

Working with our industrial partner, PTV, we undertake a capacity-based study to investigate the congestion-reducing benefits from such an information scheme including under varying degrees of passenger uptake. Our approach relies on travel-card data, from which we construct a detailed congestion model of two rail lines in the Melbourne network: Werribee and Williamstown. We combine this data with an optimisation-based model that measures the impact of passen-

Fig. 1: The south-western region of the Melbourne Metropolitan Rail Network comprising the Werribee and Williamstown lines.

ger demand shifting on congestion during times of peak demand. We show that with even modest levels of uptake (e.g., 20%) congestion measures can be almost halved. Meanwhile with an uptake of 60% we show that overcrowded trains can be almost entirely eliminated.

2 Background: The Melbourne Rail Network

The Melbourne Metropolitan Rail Network is a large hub-and-spoke system comprised of 217 stations, connected by 837km of rail [1]. Figure 1 shows the south-western region of this network, the focus of our study.

The network consists of 16 lines which connect at a central terminus, *Flinders Street Station*. Trains in the network operate from 5am to midnight on weekdays and until 1am on weekends. Morning peak demand occurs between 7:00am and 9:30am, and afternoon peak between 3:30pm and 7:00pm week days.

The network is serviced by a fleet of more than 200 trains which are managed by Metro Trains, a privately-owned rail operator. The fleet is currently made up of 3 models [1]. These are: **Comeng**, having 536–556 seats, with a target capacity of 800 passengers; and **X'Trapolis** and **Siemens**, both having 528 seats and target capacity of 900 passengers. Any train which exceeds its target capacity is considered to be in *breach* of the service agreement between PTV and Metro Trains. Breach events are undesirable because they typically cause delays at stations and increase the risk of accidents when passengers are boarding and disembarking. Systematic breaching can result in penalties for the rail operator.

3 Modelling Assumptions

3.1 Trains and Rail Network

We focus our attention on the south-western section of the Melbourne Rail Network (Figure 1). This network consists of two rail lines, with origins at Werribee and Williamstown, and three distinct types of train services:

– *Werribee Express* services, which originate at Werribee, have a frequency of approximately 10 minutes and run non-stop between the stations of Laverton and Newport and between Newport and Footscray.
– *Williamstown* services, which originate at Williamstown, have a frequency of approximately 20 minutes and stop at all stations.
– *Laverton* services, which originate at Laverton, have a frequency of approximately 20 minutes and stop at all stations, running through the so-called Altona loop.

We study demand shifting for the morning peak period. We consider all scheduled trains and we work with actual arrival times (cf. departure times) at each station, as measured by our industrial partner PTV. Additionally, owing to the configuration of the rail lines, overtaking is not possible. That means the relative order of arrival of trains at each station is fixed.

3.2 Measuring Congestion

In this section we discuss the region we study, the definitions of different levels of congestion, and the way we translate trip data into congestion measurements.

In the modelling that follows, we consider 5 levels of congestion. We use the capacity of Comeng trains as a reference, since these trains service the south-western rail network. We believe that this does not affect the generality of our conclusions. The congestion levels we use accord to PTV's own scale for congestion, and are broadly in line with those reported in [6].

1. **sparse** 0 to 264 passengers: no more than half seated capacity.
2. **seated** 265 to 528 passengers: fewer passengers than seats.
3. **standing** 529 to 662 passengers: more passengers than seats, but less than half standing capacity.
4. **target** 663 to 800 passengers: more passengers than seats, and more than half standing capacity.
5. **breach** 800+ passengers.

3.3 Data collection and train occupancy calculation

We calculate passenger counts on each train service from *smart card* data, which records the location and times at which a passenger entered and exited the rail system by *touch on* and *touch off*. The data used for this project comprised 3.6m passenger touch on and 3.5m touch off instances throughout Melbourne's rail network from January 30 to February 5 2017. We impute train occupancy from these data by assigning passenger trips to specific train services through the entry and exit stations during the corresponding time window using the following protocol. In each 10-minute period, at each station:

1. Count the passengers arriving (touch on); identify each passenger's destination (touch off location).
2. Identify *eligible services* stopping at the current station (during the current period, or immediately after if no service is available during the current period) and stopping at the touch off station.
3. Remove passengers disembarking at their touch off station.
4. Assign waiting passengers equally to eligible services.
5. Record congestion level.

We assume that city-bound passengers exiting the train system outside the south-western region remain on board the train until the last hub. For the purpose of our study, this was Footscray. Figure 2a shows train occupancy during the Thursday, February 2, 2017 morning peak period calculated using this method. (Thursday has the busiest morning peak; 15,927 trips were identified over this period.) Passenger occupancy (shaded to show congestion) highlights the overcrowding that motivated this study.

4 Greedy demand shifting

To observe the potential reduction in congestion due to passenger demand shifting, we modify the passenger load calculation to simulate passengers choosing less congested adjacent services at Laverton and Newport. This includes trains arriving up to 10 minutes earlier and departing 10 minutes after the current time period. We use greedy shifting, moving as many boarding passengers as possible in order to keep the congestion score of the current train and alternative services to a minimum. Treating levels 4 and 5 as *congested*, we use the same protocol as shown above, modified as follows:

4. Check congestion level of the incoming train; Distribute boarding passengers equally to all eligible services; for congested services, reassign a proportion of these passengers *equally* to *all* the non-congested trains arriving during the interval from 10 minutes prior, to 10 minutes post current time window.

Figure 2b shows passenger occupancy and congestion during the February 2, 2017 morning peak after greedy demand shifting is applied at Laverton and Newport. We only present results up to 8:30am to save space but note that congestion levels are low for all services departing any station in the south-western rail corridor after this time, with available seating in all cases. Comparing the two figures it is evident that the number of breach incidents decreases from 13 in the original case to 11 when greedy shifting is adopted.

Dep. Time	Werribee	Laverton	Altona	Williamstown	Newport	Footscray
7:04	435	530	-	-	707	691
7:13				171	242	238
7:16		108	256	-	499	487
7:19	532	662	-	-	900	877
7:29	620	749	-	-	971	933
7:35				202	450	433
7:38		159	358	-	578	554
7:39	648	785	-	-	1003	964
7:49	1001	1087	-	-	1233	1183
7:57				215	488	463
8:00		103	266	-	569	539
8:02	791	855	-	-	1091	1049
8:11	480	545	-	-	722	699
8:19				144	388	371
8:22		101	194	-	374	360
8:24	425	485	-	-	640	621
8:33	383	435	-	-	559	538

(a) Actual passenger count and congestion levels.

Dep. Time	Werribee	Laverton	Altona	Williamstown	Newport	Footscray
7:04	435	530	-	-	707	691
7:13				171	242	238
7:16		108	256	-	499	487
7:19	532	662	-	-	900	877
7:29	620	749	-	-	742	704
7:35				202	565	548
7:38		159	358	-	808	783
7:39	648	785	-	-	774	735
7:49	1001	961	-	-	945	895
7:57				215	684	659
8:00		229	392	-	858	828
8:02	791	762	-	-	750	709
8:11	480	592	-	-	851	828
8:19				144	471	453
8:22		148	241	-	420	407
8:24	425	485	-	-	640	621
8:33	383	435	-	-	559	538

(b) Passenger count and congestion levels with greedy demand shifting.

Fig. 2: Passenger counts and congestion levels during the morning peak for all services operating on the Werribee-Williamstown-Footscray network on February 2, 2017.

5 Reducing Congestion with Discrete Optimisation

Although the greedy approach shows some potential benefit from demand shifting, it does not reveal the greatest reduction in congestion that could be achieved if a whole network view was taken. In this section we describe a discrete optimisation model which assigns passengers to trains to reduce congestion globally. This enables us to determine the maximum reduction in congestion that could be achieved by demand shifting.

The principal data set used by the model contains trip information quantised into time blocks of 10 minutes. This shows, for each time block, the number of city-bound passengers travelling between each pair of stations during that period. From this, a flow network is constructed, which assigns the number of passengers embarking and disembarking for each service and each station. This enables the occupancy of each train to be calculated along its journey. Passengers are constrained to maintain a feasible trip across the network that is similar to their recorded trip touch on and touch off times. The objective is to minimise the congestion in the network. The optimal solution automatically reroutes passengers to reduce congestion. We recognise the solution given by this model represents an idealisation not achievable in practice, as it is based on perfect future knowledge, and assumes complete compliance by passengers. However, it does show the degree to which congestion could be reduced in a perfect situation.

The core sets of the model are given below, along with the name we shall use for indices that refer to elements in that set

ST, st	The set of stations considered
L, l	The set of lines considered
B, b	The set of time blocks considered
S, s	The set of train services considered

The core data of the problem model is given by

$mpax \in \mathbb{Z}$	Maximum passengers on any service
$seq_l \subseteq ST$	The sequence of stations for line l
$line_s \in L$	Which line is used by service s
$trip_{b,st_1,st_2} \in \mathbb{Z}$	The number of passengers commencing a journey at station st_1 in time block b to go to st_2
$comp_{b,st,s} \in \{true, false\}$	Is it possible for a passenger to enter service s at station st at time block b

The *compatibility*, $comp_{b,st,s}$, of a station st and time block b with a service s is determined as follows. Assuming we are allowing demand shifting be able to change passenger arrival times at their start station by no more than δ minutes from the time shown by recorded trip data, then s is compatible with b and st if the departure time for s from station st is no earlier than δ minutes before the time block commences, and no later than δ minutes after the time block ends. If we disallow forward shifting of passenger arrivals (i.e., assigning passengers to a train departing earlier than their touch on time), then only services that arrive between when the time block commences, and up to δ minutes later are compatible.

We assume sequence functions $first(Q)$ returning the first element q_1 of a sequence $Q = [q_1, q_2, \ldots, q_n]$, and $succ(Q)$ returning the set of adjacent pairs $\{(q_1, q_2), (q_2, q_3), \ldots, (q_{n-1}, q_n)\}$ of sequence Q. We compute auxiliary data

$on_{b,st} = \sum_{st' \in ST} trip_{b,st,st'}$	The number of passengers entering the system at station st at time b
$off_{b,st} = \sum_{st' \in ST} trip_{b,st',st}$	The number of passengers leaving the system at station st that entered at time b
$visits_{s,st} = st \in seq_{line_s}$	Whether service s visits the station st

The principle decisions *enter* and *exit* and auxiliary decision variables *pax* are defined below. They are all constrained to lie in the range $0..mpax$.

$enter_{b,st,s}$	The number of passengers entering service s at station st at time block b
$exit_{b,st_1,st_2,s}$	The number of passengers exiting service s at station st_2 that entered at station st_1 in time block b
$pax_{st,s}$	The number of passengers on service s when departing station st

We are now in a position to define the constraints of the problem.

$$enter_{b,st,s} = 0 \quad \forall b \in B, st \in ST, s \in S, \neg visits_{s,st} \tag{1}$$

$$exit_{b,st_1,st_2,s} = 0 \quad \forall b \in B, st_1, st_2 \in ST, s \in S, \neg visits_{s,st_1} \tag{2}$$

$$exit_{b,st,st,s} = 0 \quad b \in B, st \in ST, s \in S \tag{3}$$

$$pax_{st,s} = 0 \quad s \in S, \neg visits_{s,st} \tag{4}$$

Equation (1) ensures no passengers enter a service at a station it does not visit. Similarly, equation (2) ensures no passengers exiting a service commence at a station it does not visit. Equation (3) ensures no one enters and exits a service at the same station. Equation (4) ensures that the passengers for a station st not visited by a service s is 0.

$$\sum_{b \in B, s \in S} enter_{b,st,s} = \sum_{b \in B} on_{b,st} \quad \forall st \in ST \tag{5}$$

$$\sum_{b \in B, st_1 \in ST, s \in S} exit_{b,st_1,st_2,s} = \sum_{b \in B} off_{b,st_2} \quad \forall st_2 \in ST \tag{6}$$

$$\sum_{s \in S, comp_{b,st,s}} enter_{b,st,s} = on_{b,st} \quad \forall b \in B, st \in ST \tag{7}$$

$$\sum_{s \in S} exit_{b,st_1,st_2,s} = trip_{b,st_1,st_2} \tag{8}$$

$$exit_{b,st_1,st_2,s} \leq enter_{b,st_1,s} \quad \forall b \in B, st_1, st_2 \in ST, s \in S \tag{9}$$

$$pax_{st,s} = \sum_{b \in B} enter_{b,st,s} \quad s \in S, st = first(seq_{line_s}) \tag{10}$$

$$\begin{aligned} pax_{st_2,s} = pax_{st_1,s} + \textstyle\sum_{b \in B} enter_{b,st_2,s} - \textstyle\sum_{st_3 \in ST, b \in B} exit_{b,st_3,st_2,s} \\ \forall s \in S, (st_1, st_2) \in succ(seq_{line_s}) \end{aligned} \tag{11}$$

Equation (5) ensures that all passengers arriving in the system enter a train. Equation (6) similarly ensures that all passengers leaving the system at a station st_2 beginning at st_1 are exiting a service. Equation (7) ensures that every passenger entering the system at a station gets on a compatible service. Equation (8) ensures that the number of trips from station st_1 to st_2 commencing in block b matches the travel data. Equation (9) ensures that no more passengers take a trip from st_1 to st_2 on service s commencing in block b than enter the service.

Equation (10) ensures the passengers on the service s at its starting station are correct. Equation (11) ensures the passengers on the service s at later stations in the line are correct by adding in newly entering passengers and removing exiting passengers.

Finally we can specify the objective for optimising. We consider two objectives. The first is based on PTV's own congestion scale. It just counts the number of stations and services which reach each capacity level, and penalises each capacity level by a rapidly increasing amount. Let $sparse = 264$, $seated = 528$, $standing = 662$ and $target = 800$, then the first objective is simply:

$$\text{MINIMIZE} \begin{array}{l} \sum_{st \in ST, s \in S, pax_{st,s} > sparse, pax_{st,s} \leq seated} 10 \\ + \sum_{st \in ST, s \in S, pax_{st,s} > seated, pax_{st,s} \leq standing} 100 \\ + \sum_{st \in ST, s \in S, pax_{st,s} > standing, pax_{st,s} \leq target} 1000 \\ + \sum_{st \in ST, s \in S, pax_{st,s} > target} 10000 \end{array} \tag{12}$$

The objective above is deceptive: a train running at $seated + 1$ passengers is given the same objective cost as one running at *standing*. For example, this means that once a train needs more than *seated* passengers the objective will try to fill it to *standing*.

An alternative objective builds a continuous piecewise linear function which defines a cost for each passenger load, which grows steadily with higher capacities increasing faster. The function we use is

$$cost(p) = \begin{array}{ll} 0 & p \leq sparse \\ p - sparse & p > sparse, p \leq seated \\ 5 \times (p - seated) + cost(seated) & p > seated, p \leq standing \\ 10 \times (p - standing) + cost(standing) & p > standing, p \leq target \\ 100 \times (p - target) + cost(target) & p > target \end{array}$$

where $cost(level)$ represents the cumulative costs per passenger before reaching the current congestion level. Then the objective is simply

$$\text{MINIMIZE} \sum_{st \in ST, s \in S} cost(pax_{st,s}) \tag{13}$$

Restricting Passenger Movement The model as defined above allows all passengers to be shifted from their original service. While this provides a strong lower bound on possible congestion, we are unlikely to be able to enforce this behaviour. We also consider cases where only some percentage p of the customers can be moved. This reflects an assumption that any take-up in advice will only ever be followed by at most $p\%$ of customers. Adding this to the model simply requires adding lower bounds to the $enter_{b,st,s}$ variables to be $(100 - p)\%$ of the baseline ridership on each service (as computed in Section 3.3).

6 Experiments and Results

6.1 Design of Experiments

In the experiments we compare the raw congestion values determined by the train occupancy calculation of Section 3.3, with the congestion values where we enact greedy policies that divert passengers away from congested services, as well as against the discrete optimisation model of Section 5.

The discrete optimisation model is written in MiniZinc [11] and solved with the Gurobi 8.1.0 mixed integer programming solver [4]. Note that the entire model of Section 5 is linear, except for the piece-wise linear objectives. We rely on MiniZinc's automatic linearisation to encode the objective for Gurobi.

We consider experiments where we are allowed to shift 100% of passengers, which gives us a lower bound on possible congestion. To be more realistic we also consider where at most some smaller percentage $p\%$ of customers can have their behaviour changed. We examine the cases where $p = 20$, 40, 60, and 80. With $p = 0$ there is no shifting possible, we just show the calculated congestion levels.

Forward shifting of passengers, which requires some way of informing passengers to arrive earlier at the station, is more complex than simply backward shifting, which just requires information available at the station. Our experiments consider allowing both backward and forward shifting of passengers, as well as disallowing forward shifting.

To reduce the computational complexity of the optimisation model we also simplify the network by merging passenger data for stations where there is no potential for demand shifting. Therefore, Werribee incorporates Hoppers Crossing, Williams Landing and Aircraft; Williamstown incorporates Williamstown Beach and North Williamstown; Newport incorporates Spotswood, Yarraville and Seddon. Passenger travel from Westona, Seaholme and Altona is low and we consolidate demand in the so-called Altona loop.

The resulting network consists of 7 stations, with an observed daily demand ranging between 15,000 and 17,000 passengers. The network is serviced by 34 trains running during the peak period: 16 Werribee Express services, and 9 trains for each Laverton and Williamstown lines. The resulting reduced-network model requires up to 15 seconds to be solved for most instances, with a few exceptions requiring up to a maximum of 3 hours of execution.

6.2 Heatmaps

We re-examine the morning peak period for Thursday, February 2, 2017, now using the optimisation model. The resulting passenger counts and congestion levels are shown in Figure 3a assuming forward shifting is not allowed, and Figure 3b assuming forward shifting is allowed.

Clearly the optimisation-based solution drastically reduces congestion levels. With forward shifting, it is able to restrict congestion levels on all services to *standing*. Even without forward shifting, it is able to remove all breach events from the system.

Dep. Time	Werribee	Laverton	Altona	Williamstown	Newport	Footscray
7:04	505	528	-	-	662	648
7:13				171	264	261
7:16		115	264	-	662	652
7:19	510	528	-	-	662	647
7:29	614	625	-	-	739	662
7:35				202	662	654
7:38		329	528	-	662	648
7:39	662	662	-	-	662	662
7:49	762	744	-	-	759	662
7:57				215	662	649
8:00		374	528	-	798	783
8:02	542	528	-	-	662	596
8:11	516	528	-	-	662	651
8:19				144	528	521
8:22		220	338	-	528	520
8:24	454	528	-	-	525	510
8:33	452	440	-	-	528	506

(a) Forward shifting not allowed.

Dep. Time	Werribee	Laverton	Altona	Williamstown	Newport	Footscray
7:04	527	528	-	-	528	527
7:13				171	498	494
7:16		99	250	-	662	654
7:19	450	528	-	-	633	528
7:29	528	528	-	-	662	655
7:35				202	662	654
7:38		260	459	-	528	514
7:39	528	528	-	-	662	660
7:49	528	528	-	-	636	528
7:57				215	662	649
8:00		377	528	-	528	513
8:02	577	528	-	-	528	493
8:11	488	528	-	-	528	515
8:19				144	662	646
8:22		226	345	-	528	520
8:24	436	528	-	-	528	512
8:33	424	395	-	-	528	519

(b) Forward shifting allowed.

Fig. 3: Optimised passenger counts and congestion levels during the morning peak for all services operating on the Werribee-Williamstown-Footscray network on February 2, 2017.

6.3 Effect of Passenger Uptake

In the next experiment we vary passenger uptake levels, thus restricting the number of passengers that can be moved from their original service. We also consider the five different weekdays from Monday, January 30, 2017 until Friday, February 3, 2017.

Figure 4 shows on the left how the objective function of Equation (12) changes for different levels of passenger uptake, for all five weekdays. We compare greedy demand shifting versus the optimisation-based demand shifting using the objective of Equation (12), with and without forward shifting. Note that Monday has noticeably less passengers than the other days. Clearly, greedy demand shifting

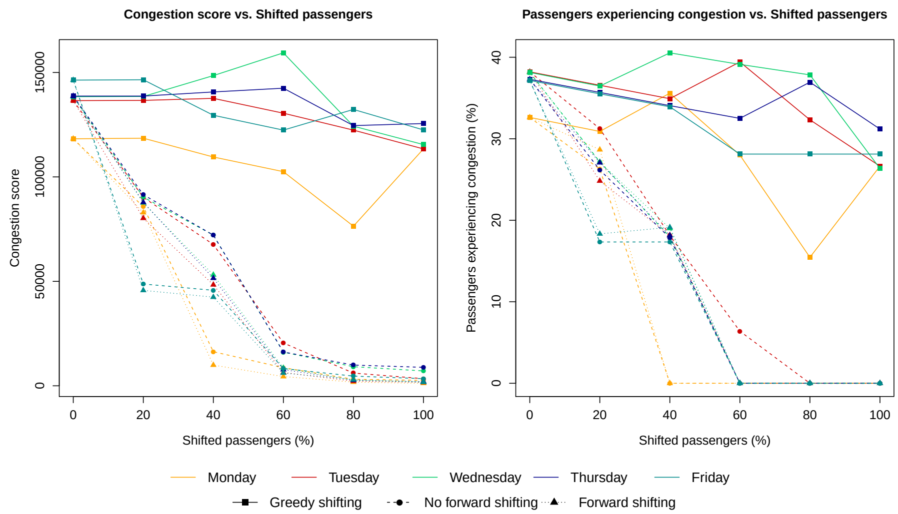

Fig. 4: Congestion levels (left) and percentage of passengers experiencing congestion at some point in their trip (right), using Equation (12).

only has a slightly beneficial effect on the network and can indeed *worsen* the congestion score, because it makes myopic decisions which end up leading to later congestion. In contrast, the optimisation-based approaches can drastically reduce congestion. With 100% take-up, congestion is always reduced to nearly zero when allowing forward shifting. Enabling forward shifting, while universally improving the results, does not appear to make that much difference, at least at the granularity visible in the plot. But the gains are substantial when considered in relative terms (c.f. Figure 3).

Figure 4 on the right shows how many passengers experience congestion on their trip, that is, passengers that travel at least one segment in a breached train. Again we see the greedy demand shifting can worsen this measure, while the optimisation approaches can quickly find solutions where no passenger experiences congestion.

Figure 5 shows the results measured with the more fine-grained objective function of Equation (13). In these experiments, the optimisation approach was run minimising this objective. On the left we see how the objective value changes as passenger uptake increases. Interestingly, using this measure we can see that greedy shifting is in fact reducing congestion per passenger, although not per service, and it does improve as passenger uptake increases. The optimisation solutions are again far superior, reducing the objective function to very low values smoothly as the uptake increases.

Figure 5 on the right shows the percentage of passengers experiencing congestion. The results for greedy shifting are unchanged. Again, the optimisation results show a significant reduction on the proportion of passengers travelling on

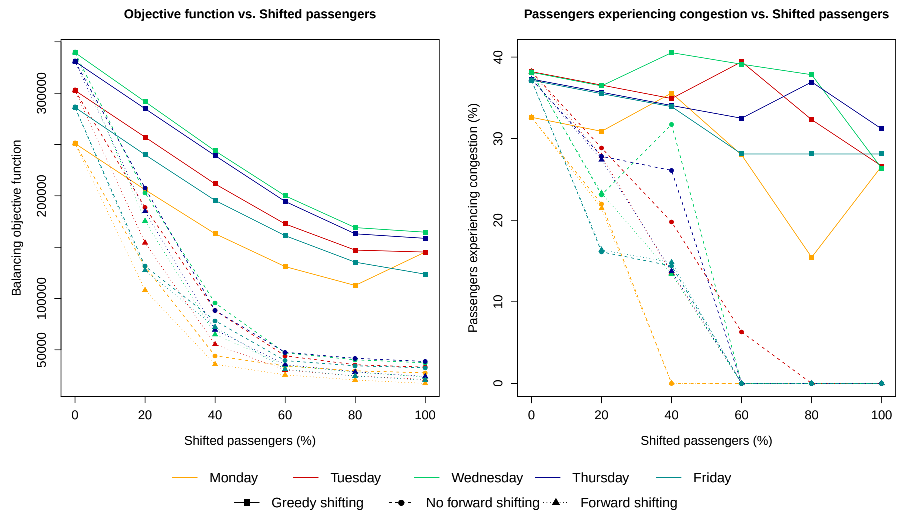

Fig. 5: Balancing objective function values (left) and percentage of passengers experiencing congestion at some point in their trip (right), using Equation (13).

breached services. The peak for Wednesday and 40% uptake can be explained by considering that our model is not directly minimising the percentage of passengers experiencing congestion. According to the recorded trips, several services ran well above the breach threshold on this day. Considering a small uptake (e.g., 20%) evens out the number of passengers on these services, but keeps most trains previously running at *target* on the same congestion level. At 40% uptake, the number of shifted passengers is not enough to bring the former services below *breach*, but some passengers can be re-allocated to different trains that were operating just under the breach threshold. Passengers on these services, who were not considered to be experiencing congestion before, are now travelling on breached trains. However, despite the percentage increasing, passenger numbers on board over-congested trains are lower and more balanced across services. Table 1 summarises the results obtained with the proposed fine-grained objective function aimed at balancing passengers between services, including the number of breach incidents and the percentage of passengers experiencing congestion for different levels of take-up.

6.4 Experimental Results and Discussion

If forward shifting is not allowed, the model yields results that will help alleviate congestion, but do not completely eliminate it. We still obtain some trains where utilisation is very close to the *breach* threshold. This increases the risk of over-congestion if passenger numbers keep rising. However, this kind of intervention is the easiest and most likely the cheapest, since it does not require a major

Table 1: Results obtained with our optimisation model using Equation (13) for different percentages of shifted passengers showing number of breach incidents and percentage of passengers experiencing a breach service.

		0%		20%		40%		60%		80%		100%	
Day	Forward	Breach inc.	PAX (%)	Breach inc.	PAX (%)	Breach inc.	PAX (%)	Breach inc.	PAX (%)	Breach inc.	PAX (%)	Breach inc.	PAX (%)
Mon	×	11	32.62	8	22.02	0	0.00	0	0.00	0	0.00	0	0.00
	✓	11	32.62	8	21.45	0	0.00	0	0.00	0	0.00	0	0.00
Tue	×	13	38.21	10	28.86	7	19.79	1	6.30	0	0.00	0	0.00
	✓	13	38.21	9	27.84	4	13.53	0	0.00	0	0.00	0	0.00
Wed	×	13	38.13	9	23.11	10	31.75	0	0.00	0	0.00	0	0.00
	✓	13	38.13	9	23.32	4	13.42	0	0.00	0	0.00	0	0.00
Thu	×	13	37.29	10	27.83	7	26.11	0	0.00	0	0.00	0	0.00
	✓	13	37.29	10	27.42	4	13.76	0	0.00	0	0.00	0	0.00
Fri	×	14	37.16	4	16.12	4	14.40	0	0.00	0	0.00	0	0.00
	✓	14	37.16	4	16.36	4	14.84	0	0.00	0	0.00	0	0.00

change in passenger behaviour. Arrival and touch-on patterns remain the same, but passengers are given recommendations to board specific alternative trains in order to avoid an uncomfortable trip on a congested service. This could be achieved by, e.g., providing additional on-screen information about alternative trains and congestion levels. One example of this is the smartphone app of NS, the Dutch national rail operator, which shows the expected congestion levels of arriving trains, from which passengers can make an informed choice of whether to wait for a less congested service or not. We observe the greatest reduction in congestion is achieved when forward shifting is allowed. For this to work, passengers might need to arrive at the station up to 10 minutes prior to their intended trip departure for a less congested ride. This would require a change in passenger behaviour, and might need to be implemented in conjunction with other incentive mechanisms, such as fare reduction. However, our results show that even a quite modest adoption of such a program (of the order of 20%) could provide a significant reduction on congestion levels during the morning peak.

Our optimisation results also show that demand shifting up the line (closer to where the service originates) can lead to a major reduction in congestion down the line—that is, closer to Melbourne. The results for simulated demand shifting show that myopic interventions lead at most to a minor alleviation of congestion for some services. The implications of this research for PTV is that better management of demand originating at Werribee, Hoppers Crossing, Williams Landing, or Aircraft, could reduce congestion at busy stations down the line. For example, many trains originating at Werribee currently depart Laverton and Newport at or above *target* levels, so that it is almost inevitable that these trains will become congested as they journey towards Melbourne. Results from our optimisation model show that reducing demand for these services when alternatives are available at origin could prevent congestion down the line. Our results, thus, give some guidance on where to trial interventions to reduce congestion in the morning peak.

7 Conclusions and Future Work

Much has been written on the various ways in which to optimise rail networks, but relatively little work that we are aware of exists on providing information for passengers to modify their behaviour in ways that improve the system for all users. Our pilot study shows how this might be achieved in the southwest Melbourne rail network. As Melbourne grows, the efficient usage of all public transport infrastructure will become more and more important, and modifying passenger behaviour is an attractive alternative to provisioning more services.

There are several ways in which the work presented here might be extended. The actual weights (or costs) in our objective functions from Section 5 are ad hoc. Analysis of the relative importance ascribed to comfort over the other competing concerns of passengers from the survey literature may, perhaps, more realistically weight the objective function in our model. We could consider the multi-objective problem measuring both trip time and comfort and explore the efficient (Pareto) set of possible solutions arising from this.

Deeper analysis of customer arrival times at stations could also be valuable. Patterns in customer arrival may arise from a mixture of behaviours—for example, we would expect to see *schedule-aware* customers whose arrival time is some function of when the next train is scheduled to depart, and *schedule-oblivious* customers whose arrival time is largely independent of the timetable. This leads to the possibility of using mixture modelling [10], possibly incorporating the Poisson distribution [15], to more accurately model customer behaviour. Further to such mixture modelling for different customer arrival times, re-visiting the discussion of [17], our greedy and optimisation models could be modified to treat customers of various behaviour types differently.

8 Acknowledgements

We acknowledge the Monash University Faculty of Information Technology for seed funding to begin this project. Daniel Harabor is funded by the Australian Research Council under the grant DE160100007.

References

1. Network Development Plan - Metropolitan Rail December 2012 (Updated 2016). Public Transport Victoria (2016)
2. Annual Report 2017-18. Public Transport Victoria (2018), 140pp
3. Frappier, A., Morency, C., Trépanier, M.: A new method to measure the quality and diversity of transit trip alternatives. Tech. rep. (October 2015), CIRRELT (Centre interuniversitaire de recherche sur les réseaux de l'interprise, la logistique et le transport) technical report CIRRELT-2015-51
4. Gurobi optimizer version 8.1. http://www.gurobi.com/products/gurobi-optimizer (2019)

5. Haywood, L., Koning, M.: The distribution of crowding costs in public transport: New evidence from Paris. Transportation Research Part A: Policy and Practice **77**, 182–201 (2015)
6. Kroes, E., Kouwenhoven, M., Debrincat, L., Pauget, N.: On the value of crowding in public transport for Ile-de-France. international transport forum discussion papers, no. 2013/18 (2013)
7. Kroes, E., Kouwenhoven, M., Duchateau, H., Debrincat, L., Goldberg, J.: Value of punctuality on suburban trains to and from paris. Transportation Research Record **2006**(1), 67–75 (2007)
8. Liu, R., Li, S., Yang, L.: Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy. Omega (2018)
9. Liu, Y., Charles, P.: Spreading peak demand for urban rail transit through differential fare policy: a review of empirical evidence. In: Australasian Transport Research Forum 2013. Queensland University of Technology, Brisbane, QLD (2013)
10. McLachlan, G.J., Peel, D.: Finite mixture models. Wiley (2000)
11. Nethercote, N., Stuckey, P.J., Becket, R., Brand, S., Duck, G.J., Tack, G.: MiniZinc: Towards a standard CP modelling language. In: Bessière, C. (ed.) Principles and Practice of Constraint Programming – CP 2007. LNCS, vol. 4741, pp. 529–543. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
12. de Palma, A., Lindsey, R., Monchambert, G.: The economics of crowding in rail transit. Journal of Urban Economics **101**, 106–122 (2017)
13. Pluntke, C., Prabhakar, B.: INSINC: a platform for managing peak demand in public transit. JOURNEYS, Land Transport Authority Academy of Singapore pp. 31–39 (2013)
14. Shi, J., Yang, L., Yang, J., Gao, Z.: Service-oriented train timetabling with collaborative passenger flow control on an oversaturated metro line: An integer linear optimization approach. Transportation Research Part B: Methodological **110**, 26–59 (2018)
15. Wallace, C.S., Dowe, D.L.: MML clustering of multi-state, Poisson, von Mises circular and Gaussian distributions. Statistics and Computing **10**(1), 73–83 (2000)
16. Yen, B.T., Tseng, W.C., Chiou, Y.C., Lan, L.W., Mulley, C., Burke, M.: Effects of two fare policies on public transport travel behaviour: Evidence from South East Queensland, Australia. Journal of the Eastern Asia Society for Transportation Studies **11**, 425–443 (2015)
17. Zhang, J., Yang, H., Lindsey, R., Li, X.: Modeling and managing congested transit service with heterogeneous users under monopoly. Transportation Research Part B: Methodological (2019)