OPTIMAL INFRASTRUCTURE CAPACITY FOR RAIL-BOUND DEMAND RESPONSIVE TRANSIT

Oded Cats, Corresponding author
Department of Transport and Planning
Delft University of Technology
P.O. Box 5048, 2600 GA Delft, The Netherlands
Phone number: +31 15 2781384
Fax number: +31 15 2787956
Email: o.cats@tudelft.nl

Jesper Haverkamp
Nederlandse Spoorwegen
Department of Operations Control, Section Business Intelligence
P.O. Box 2025, 3500 HA Utrecht, The Netherlands
Phone number: +31 6 1333 7136
Email: jesper.haverkamp@ns.nl
AND
Department of Transport and Planning
Delft University of Technology

Paper submitted for presentation at the 97th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2018 and publication in the Transportation Research Record, Journal of the Transportation Research Board

Word Count: c.a. 5730 + 2 Tables (500) + 5 Figures (1250) = 7470 words

Keywords: Automated Vehicle, Demand Responsive, Rail Infrastructure, Railway Planning, Capacity Allocation.
ABSTRACT

Fully-automated services allow for greater flexibility in operations and lower marginal operational costs. The objective of this study is to determine the capacity requirements of an envisaged automated rail demand responsive transit (DRT) system which offers a direct non-stop service. An optimization model for determining the optimal track and station platform capacities for a rail-DRT system so that passenger, infrastructure and operational costs are minimized is formulated. The macroscopic model allows for studying the underlying relations between technological, operational and demand parameters, optimal capacity settings and the obtained cost components. The model is applied to a series of numerical experiments followed by its application to part of the Dutch railway network. The results of the numerical experiments and the case study application indicate that - unlike conventional rail systems in which stations often are capacity bottlenecks - link capacity properties are more critical for the performance of automated rail-DRT systems than station capacity. The performance is benchmarked against the existing service suggesting that in-vehicle times can be reduced by 10% in the case study network with the optimal link and station capacity allocation comparable to those currently available in this heavy rail network. A series of sensitivity analyses was performed to test the consequences of network and demand settings as well as the characteristics of future automated rail-DRT systems.
INTRODUCTION
The rapid advancements in the developments of fully-automated vehicles has led to an increasing interest in the concept of demand responsive transit (DRT) systems. Automated services will allow for greater flexibility in operations and lower marginal operational costs. The area of application has almost exclusively been limited to road-bound systems. To the best of the authors knowledge, applying DRT as a substitute for current heavy rail services has not been considered in the literature or practice insofar. This paper presents a first step into the relatively unknown area of rail DRT.

In this work, a DRT system that is designed as a full replacement of scheduled heavy rail for a given (sub-)network is envisaged. The automated rail-bound vehicles offer a direct non-stop service and move in the rail network in response to passenger requests with no predefined routes and schedules. Vehicles transport passengers that share the same origin and destination stations. Vehicles can be sized according to the operator’s preference, but they are considerably smaller than existing trains.

The objective of this study is to determine the capacity requirements of rail DRT. Unlike road-bound DRT services operate in rural areas or cater for special user groups, the rail-DRT is designed to serve a large geographical area with relatively high demand that can result in operation constrained by congestion and capacity limitations. The strategic planning objective of this study constitutes a major difference from most earlier models such as Anderson [1] and Winter et al. [2] which considered microscopic, operational models with a stochastic passenger arrival.

The development of technological and service concepts that will facilitate rail-DRT systems are still at their early stages. It is therefore not surprising that literature on rail-DRT is limited. An early research identified the challenges of short vehicle headways and limited station capacity in the context of dense urban operations [3]. The vehicle engineering RailCab project developed technical and mechanical solutions for small driverless rail-bound traffic [4]. Vehicle design solutions for operating at short headways in an automated guideway transit system (AGT) were studied by Choromanski and Kowara [5], while capacity in relation to station layout has been analysed in more detail by Greenwood et al. [6]. While these studies provide preliminary insights into anticipated advanced in-vehicle technology, there is lack of knowledge on the capacity requirements that such operations inflict on railway network infrastructure and related system performance and level-of-service.

This paper presents a novel optimization model for determining the optimal track and station platform capacities for a rail-DRT system so that passenger, infrastructure and operational costs are minimized. The model is formulated as a cost minimization problem with the premise that system-optimum vehicle flow distribution conditions can be attained. The model allows for studying the underlying relations between technological, operational and demand parameters, optimal capacity settings and the obtained cost components. The model is applied to a series of numerical experiments followed by its application to part of the Dutch railway network. The performance is benchmarked against the existing service and its sensitivity to various scenarios is assessed and the implications of which are discussed.

METHODOLOGY
Modelling Approach
Current railway models with timetables as their cornerstone are unsuitable for rail DRT applications. The main challenge in modelling network-wide long-term planning for rail DRT is that the model needs to capture flow distribution and capacity constraints without representing system dynamics in microscopic details. Other rail DRT challenges include the need to handle large numbers of hourly passenger requests, strict routing constraints due to the nature of rail infrastructure and highly heterogeneous service characteristics compared to
traditional rail systems. The approach taken in this study is to develop a novel macroscopic model by considering rail DRT as a special case of the network flow problem. This approach allows evaluating a large number of network-wide solutions and performing an exact optimization that guarantees to attain a global optimum.

The deterministic and static optimization problem is solved for a given network topology and passenger demand distribution, both are thus exogenous to the model. Other input to the model include vehicle size (homogenous fleet), cost units and track and node flow-speed and -delay functions. Model output are the optimal infrastructure capacity per network element (i.e. rail segments and stations), vehicle flow distribution and the value of passenger, infrastructure and operational costs.

### Network Representation and Cost Functions

The public transport network is represented by a directed and weighted graph $G(N, A)$, where the node set $N$ represents rail stations, and the set of links $A \subseteq N \times N$ represents rail track segments between stations. $s_j$ is the capacity of station $j \in N$. $c_a$ is the capacity of link $a \in A$ and $L_a$ is the corresponding length. $a_{i,j}$ denotes the directed link connecting station $i$ to downstream station $j$. Travel demand is given in the form of an hourly OD matrix, $P$, where each entry, $P_{od}$, denotes passenger demand between a pair of stops $o$ and $d$, $o, d \in N$. Let $x_{a,od}$ represent the flow on link $a$ that travels between a certain OD pair. Link speed-density function and node delay-density function are detailed in the following sections.

#### Link travel times

The speed-density function is assumed to follow a logistic function [7] albeit with an abrupt transition from free-flow to jam which is more suitable for automated systems. Let $v_a$ and $\hat{v}_a$ be the speed and the free-flow speed, respectively, of vehicles traversing link $a$. Link travel times are then determined based on link length, free-flow speed and the logistic term with the volume over capacity ratio as follows:

$$t_a = \frac{l_a}{\hat{v}_a \left[ 1 + e^{-\alpha \left( \frac{x_a - \varphi}{\hat{v}_a} \right)} \right]}$$

(1)

The parameters $\alpha$ and $\varphi$ are the scale and shift of the logistic distribution. $\theta$ is the maximum number of vehicles per hour per unit of link capacity to convert infrastructure units (i.e. number of tracks) into flow units (i.e. number of vehicles).

#### Station processing times

Let $x_{j,od}$ denote the number of vehicles that arrive at station $j$ that are en-route connecting a given OD-pair (including vehicles for which this station is an origin or a destination). Vehicle arrivals are assumed to follow the Poisson distribution, implying that service requests of all vehicles can be represented as a joint Poisson process with the summative event rate parameter $\lambda_j = \sum_{o \in N} \sum_{d \in N} x_{j,od}$.

Vehicles either drive through or call at origin and destination stations. Considering each platform as a server and assuming that all vehicles have the same mean service time $1/\mu_j$ with an exponential service time distribution and all vehicles can use all platforms, the DRT station is characterized as a non-pre-emptive M/M/c system (similarly to metro stations in for example [8]). If the station has more than two platforms, it is assumed that through-going vehicles can overtake dwelling vehicles, otherwise the station is governed by non-prioritized M/M/c queues. The value of $\mu_j$ is determined based on the ratio between through-
going and dwelling vehicle-times at each station and thus depends on the share of vehicle
calling (i.e. originating or destined) at station $j$ out of all vehicles traversing this stations and
dwell time, $\tau_{dwell}$. The expected waiting time in the non-pre-emptive M/M/c queue
depends on delay probability $P$ [9], which is calculated as follows:

$$P_j = \frac{(s_f \rho_j)^{s_j}}{s_f} \cdot \frac{1}{(s_f \rho_j)^{s_j} + (1 - \rho_j)} \sum_{m=0}^{s_j-1} \frac{(s_f \rho_j)^m}{m!} \quad \forall j \in N$$ (2)

Where the density term is defined as:

$$\rho_j = \frac{\Sigma_{o \in N \setminus j} \Sigma_{d \in N \setminus j} x_{j,od}}{s_f \mu_j} + \frac{\Sigma_{o \in N \setminus j} x_{j,od} + \Sigma_{d \in N \setminus j} x_{j,jd}}{s_f \mu_j} \quad \forall j \in N$$ (3)

The nominators in the first and second terms in Eq. 3 correspond to the number of vehicles traversing and calling, respectively, at station $j$. The corresponding expected delays in prioritized and non-prioritized queuing systems can then be determined according to the formulations provided in Wagner [10], resulting in the expected delay per station, $E(w_j) = f(P_j)$. This function assigns different values for through-going vehicles and dwelling vehicles at stations where overtaking is possible and otherwise no distinction is made.

**Cost Minimization Problem Formulation**

Considering rail DRT as a special case of the network flow problem, the decision variables are link capacity, node capacity and the share of vehicle flow routed via each route alternative per Origin-Destination pair. The objective is to minimize the cumulative value of infrastructure capacity costs, passenger travel time costs and operational costs. The rail DRT planning objective is to balance between the costs of adding infrastructure capacity and the costs of delay or detours caused by insufficient infrastructure capacity.

The cost minimization problem is then formulated as follows:

$$z = \min \left[ \beta_1 \cdot \kappa \cdot \left[ \Sigma_{a \in A} \Sigma_{o \in N} \Sigma_{d \in N} (x_{a,od} \cdot t_a) + \Sigma_{j \in N} \left( \lambda_j \cdot E(w_j) \right) \right] + \beta_2 \cdot \Sigma_{a \in A} c_a \cdot l_a + \beta_3 \cdot \Sigma_{j \in N} s_j + \beta_4 \cdot \kappa \cdot \Sigma_{o \in S} \Sigma_{d \in S} \Sigma_{a \in A} (x_{a,od} \cdot l_a) \right]$$ (4)

Subject to

$$\Sigma_{i \in N} x_{a_{i-j,od}} = \Sigma_{k \in N \setminus o,d} x_{a_{j-k,od}} \quad \forall o,d \in N$$ (5)

$$\Sigma_{j \in N} x_{a_{o-j,od}} = \frac{p_{od}}{\kappa} \quad \forall o,d \in N$$ (6)

$$\Sigma_{j \in N} x_{a_{j-d,od}} = \frac{p_{od}}{\kappa} \quad \forall o,d \in N$$ (7)

$$x_{a,od} \geq 0 \quad \forall a \in A, \forall o,d \in N$$ (8)

$$c_a \geq 0 \quad \forall a \in A$$ (9)

$$s_j \in \mathbb{Z}^+ \quad \forall j \in N$$ (10)
\[ \rho_j \leq 1.0 \quad \forall j \in N \]  

In addition, Eq. 1-3 given in the previous section describe how \( t_a \), link travel time, and \( E(w_j) \), the expected delay at station, are calculated. The remaining constraints pertain to demand satisfaction, flow conservation and non-negative decision variables. Eq. 5 ensures flow conservation at all intermediate nodes whereas Eq. 6-7 require demand satisfaction at origins and destinations, respectively. Eq. 8-9 require that link flows and capacities are non-negative. Eq. 10 ensures that station capacity is a positive integer and Eq. 11 determines the upper value of station density. Decision variables are \( x_{a,od}, c_a \) and \( s_j \). \( \kappa \) is the desired vehicle load level (e.g. seat capacity).

The objective function, Eq. 4, consists of: (i) passenger travel times calculated over all links and stations; (ii) track capacity investment cost; (iii) station capacity investment cost; (iv) variable operational costs. \( \beta_1 \ldots \beta_5 \) are the monetary costs associated with each of the objective function components: \( \beta_1 \) is passenger value-of-time, \( \beta_2 \) and \( \beta_3 \) are the costs of track infrastructure and station platform capacity units expressed in hourly terms, respectively, \( \beta_4 \) corresponds to the operational costs per seat kilometre. A single arrival rate per station appears in cost component (iii) for simplicity, but different event rates are specified in case of different vehicle classes (i.e. through-going and dwelling) need to be considered.

Passenger travel times consist of waiting time and in-vehicle time in the envisaged context of a direct rail DRT service. Given the on-demand character of the system, waiting times can be approximated as half the interval between departures of vehicles connecting a given OD-pair. Service frequency is specified in this study as approximately proportional to passenger demand for a given OD pair by rounding the ratio between \( P_{od} \) and the desired vehicle load level, \( \kappa \), unless passenger demand is not sufficient to justify a predefined minimum service frequency. Hence, waiting times are determined in the initialization phase, independent of the decision variables and can be left out in the analysis of alternative solutions.

The minimization problem formulated in Eq. 4 entails the simultaneous solution of setting the capacity per rail-segment and station and obtaining the corresponding system-optimum solution of network flow distribution while minimizing user, investment and operational costs.

**Model Implementation and Specification**

Model specification involves setting values for a series of technological and service parameters. In the following, base case values designed to reflect the prevalent conditions in the Netherlands - for which the authors have access to relevant information and the model is later applied - are specified.

Objective function weights were set to \( \{ \beta_1, \beta_2, \beta_3, \beta_4 \} = \{10,50,123.5,0,02 \} \). The first weight is the Dutch value-of-time expressed in Euros per hour. The second and third are specified based on estimates in the Dutch rail industry and correspond to €50 million per kilometre of track and €123.5 million per station platform assuming a 38 meter long platform required for the envisaged system operations. Both are assumed to be depreciated over a 30 years period. Operational cost per seat-km is estimated at €0.02.

The speed-density function involves specifying the free-flow speed and track vehicle capacity. The former was set to 100 kilometres per hour (\( \hat{v}_a = 100 \quad \forall a \in A \)) based on current conditions in the dense Dutch network. Track capacity, \( \theta \), is set to 180 vehicles per hour in line with operational people mover systems which operate at 20 second headways. The logistic function scaling and shifting parameters, \( \alpha \) and \( \varphi \), are set to -11.17 and 0.88, respectively, based on engineering assumptions derived from a combination of the EU
publication on railway capacity utilization [7] and literature on traffic flow properties of
automated vehicles (see review in [11]).

Station platform operation is governed by the queuing servers. Based on observations
in an existing automated system (i.e. Rivium Parkshuttle) dwell time, $\tau_{\text{dwell}}$, is set to 20
seconds. Mean service rate, $1/\mu$, is set to 5 seconds for non-stop vehicles and 80 seconds for
dwelling vehicles. The latter is based on an estimation of the time required for setting
switches, pulling into the station, dwelling and clearing the platform.

Finally, the base case vehicle capacity is set to $\kappa = 24$ passengers. Vehicles are
designed so that all passengers are seated, Service frequency per OD-pair is determined to
satisfy the demand with a load factor of at least 0.7. OD-pairs for which demand does not
justify at least one departure per hour (i.e. 17 passengers) remain unserved by the DRT
service.

The optimization problem is solved in MATLAB. To reduce the number of decision
variables for real-size problems, $x_{a,\text{od}}$ was replaced with the determination of passenger flows
per OD-pair at the route rather than link level. For each OD-pair, all routing alternatives that
involve free-flow travel times of up to three times those of the shortest path were included.
Flow conservation constraints are then removed and a constraint ensuring the cumulative
route flows correspond to passenger demand per OD-pair are added instead.

NUMERICAL EXPERIMENTS

The cost minimization model described in the previous section was applied to a series of
umerical experiments to analyse the generic properties of rail DRT systems, underlying
relations between model variables and the sensitivity of model outputs to variations in input
parameters.

Experimental Set-up

The numerical experiments are performed using a graph composed of 17 nodes and 48
unidirectional links. Two distinct network structures are considered: a grid and a ring/radial
structure, as illustrated in Figure 1. The base case passenger demand corresponds to 34,000
hourly requests distributed over the network based on a gravity model using the Euclidean
distance between all origin and destination nodes (an average of 125 trips per OD-pair). Each
optimization problem is solved within 15 seconds on a standard PC.

A series of sensitivity analyses pertaining to service design, cost parameters,
technological capabilities and demand scenarios are performed. In the following the results
for variations in vehicle passenger capacity, track capacity and demand distribution pattern
are reported and discussed. In the case of vehicle capacity, it is also varied in conjunction with
 corresponding changes in operational costs and minimum service frequency threshold.

Figure 1: The two network structures considered in the numerical experiments: grid
(left) and ring/radial (right). All lines represent bidirectional arcs.

Results and Analysis
Table 1 summarizes the results of the base case scenario along with the key sensitivity analysis scenarios. The base case is set with the parameter values specified in the Model Implementation and Specification section – including a vehicle capacity of 24 passengers and track capacity of 180 vehicles per hour – and with the Euclidian gravity demand distribution. The table reports results for other vehicle capacity (VC) and track capacity (TC) scenarios noted with the corresponding value, as well as demand distribution (DD) scenarios as detailed below. Table 1 includes the four cost components for the optimal solution expressed in hourly terms (columns 2-5), followed by two indicators of system resources – fleet size and total seat-km offered – and the resulting service speed.

Overall, the grid network is found to outperform the ring/radial network, yielding lower values for all cost components in the optimal solution with the existing demand distribution. A careful investigation reveals that this difference stems primarily from the fact that in the grid network flows can often be rerouted at constant mileage, while in the ring/radial network rerouting always comes at a price of increased travel distance. Consequently, flow rerouting is common in the grid network, whereas all vehicles take the shortest route in the ring/radial scenario.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>47.1; 49.1</td>
<td>31.8; 33.5</td>
<td>46.1; 46.3</td>
<td>9.9; 10.4</td>
<td>370; 386</td>
<td>495.4; 521.6</td>
<td>74; 74</td>
</tr>
<tr>
<td>VC_12</td>
<td>58.2; 60.3</td>
<td>69.8; 72.9</td>
<td>83.3; 83.2</td>
<td>11.7; 12.2</td>
<td>915; 948</td>
<td>583.5; 609.2</td>
<td>70; 71</td>
</tr>
<tr>
<td>VC_48</td>
<td>32.4; 36.4</td>
<td>14.0; 13.5</td>
<td>27.0; 26.8</td>
<td>7.0; 7.8</td>
<td>127; 143</td>
<td>348.7; 390.1</td>
<td>75; 75</td>
</tr>
<tr>
<td>VC_96</td>
<td>26.5; 27.5</td>
<td>8.3; 7.6</td>
<td>18.1; 18.0</td>
<td>5.5; 5.9</td>
<td>52; 54</td>
<td>276.6; 294.6</td>
<td>73; 75</td>
</tr>
<tr>
<td>TC_30</td>
<td>60.6; 63.3</td>
<td>160.5; 169.0</td>
<td>46.4; 46.3</td>
<td>9.9; 10.4</td>
<td>476; 497</td>
<td>495.4; 521.6</td>
<td>57; 58</td>
</tr>
<tr>
<td>TC_45</td>
<td>55.6; 58.0</td>
<td>111.0; 116.9</td>
<td>46.4; 46.3</td>
<td>9.9; 10.4</td>
<td>437; 456</td>
<td>495.4; 521.6</td>
<td>62; 63</td>
</tr>
<tr>
<td>TC_120</td>
<td>48.7; 50.7</td>
<td>45.8; 48.2</td>
<td>46.4; 46.3</td>
<td>9.9; 10.4</td>
<td>382; 399</td>
<td>495.4; 521.6</td>
<td>71; 72</td>
</tr>
<tr>
<td>DD_CC</td>
<td>80.1; 75.2</td>
<td>57.9; 53.9</td>
<td>52.7; 52.4</td>
<td>18.0; 16.8</td>
<td>630; 591</td>
<td>902.1; 839.4</td>
<td>79; 78</td>
</tr>
<tr>
<td>DD_U</td>
<td>78.6; 74.1</td>
<td>56.9; 53.4</td>
<td>49.5; 49.5</td>
<td>17.7; 16.7</td>
<td>618; 582</td>
<td>887.1; 832.5</td>
<td>79; 79</td>
</tr>
</tbody>
</table>

VC (Vehicle Capacity); TC (Track Capacity); DD (Demand Distribution); CC (Closeness Centrality); U (Uniform)

Solution sensitivity to vehicle capacity was tested by varying it with $\kappa = \{12, 24, 48, 96\}$. These capacities correspond to the number of seats in van, base case, bus and intercity train car unit, respectively. Station capacity costs were adjusted to correspond to changing platform length requirements. A pronounced trend can be observed for both network structures in the results reported in Table 1 - the smaller the vehicle capacity the higher all cost components become. Reducing vehicle capacity from the base case value of 24 to 12 results in an increase of approximately 120% and 80% in link and node capacity costs, respectively. This increased capacity is required to handle a fleet size that is 2.5 times larger, but nevertheless results with a 25% rise in passenger costs. Part of this difference is attributed to the increasing number of passengers that are not offered a direct DRT service as vehicle size increases. This share increases dramatically from negligible levels to a sizeable minority of passenger demand in the scenario with a capacity of 96. With the exception of the lowest capacity scenario, service speed remains at a relatively constant level.

Smaller rail vehicles are associated with higher operational costs ($\beta_0$), because the seat-to-engine ratio is lower. It might therefore be argued that to ensure a fair comparison, a change in vehicle size needs to be accompanied with an opposite change in operational costs. The following scenarios are studied for the grid network: 12 seats at €0.08 per seat kilometre, 24 seats at €0.02 per seat kilometre (base case), 48 seats at €0.015 per seat kilometre and 96
seats at €0.0125 per seat kilometre. The results show that a change in unit operational costs has no major influence on the decision variables. Differences between the various vehicle size scenarios are in line with the results reported in Table 1.

To overcome the deficiency of large shares of unserved demand when vehicles size increases, the second set of combinatorial scenarios study a decrease of minimum service frequency at an increase in vehicle capacity for the grid network. The largest vehicle is associated with a frequency threshold of one departure per hour. The threshold values for smaller vehicles are set such that the minimum capacity of 96 seats per hour is obtained in all scenarios: 12 seats and at least 8 departures per hour; 24 seats and 4 hourly departures; 48 seats and 2 hourly departures; and 96 seats with at least 1 departure per hour. In this way, the share of unserved demand is fixed in all cases (12%). While served demand is equal in all scenarios, passenger costs still decrease with increasing vehicle capacity, albeit more modestly (i.e. 46,940 and 40,560 for 12 seats and 96 seats, respectively) than reported in Table 1. Link and node capacity costs are significantly higher in low vehicle capacity scenarios with costs more than five and three times as high for link and node costs, respectively, when comparing 12-seats and 96-seats scenario, yet resulting in a lower operational speed. Notwithstanding, smaller vehicles may still be preferred because they offer higher service frequency.

The automated rail track capacity assumed in the base case scenario, $\theta = 180$ vehicles per hour, is very high compared to existing heavy rail systems. In order to test the consequences of technological developments that result in lower values, the model was run with track capacity values of 30, 45 and 120 vehicles per hour. The most conservative value corresponds to the maximum frequency in classical heavy rail with ERTMS signalling. A reference value of 45 is taken from automated metro systems. An intermediate value of 120 can be indicative of more limited technological advancements. As can be expected, link costs are most affected by changes in track capacity since they exercise a linear relation. In contrast, the impact of track capacity on passenger costs exhibits a logistic relation through the operational speed. Node costs and operational costs remain unaffected. Link costs dominate the cost function for low track capacity values and decrease to levels similar to station capacity and passenger costs for track capacity values of 120 vehicles per hour and then exceeded by the latter two for $\theta = 180$ (i.e. base case).

Finally, the demand distribution is expected to have profound effects on system performance. Two demand distribution scenarios are considered in addition to the base case which has demand oriented towards the network’s centre of gravity: the opposite case of uniform demand distribution (DD_U) and an intermediate case where demand is proportional to the node closeness centrality (DD_CC) metric (i.e. average distance to all other nodes). Note that unlike the uniform and gravity distributions, demand distribution in the later scenario is not independent from network connectivity and hence results with a different OD-matrix for the two network structures. In contrast to the base case results, link costs in the optimal solutions for the two alternative demand distribution scenarios surpass station costs. A more uniformly distributed demand requires a larger fleet and leads to higher mileage and thus track capacity on more network links, albeit with lower congestion levels as reflected in the increase in average speed for both network structures. Operational costs are lower for the ring/radial network than for the grid network in the average closeness and uniform scenarios due to its better connectivity whereas the grid network performs better when demand is concentrated at the centre where it offers shorter routes.

APPLICATION
The rail-DRT capacity allocation optimization problem detailed above was applied to the real-world case study of part of the national Dutch railway network. The case study is described, followed by the results and their analysis.

**Case Study Description**

The case study area covers the railway network in the Dutch province of North-Holland, north of the North Sea Channel, shown in Figure 2. The Dutch capital, Amsterdam, is situated at the south-east corner of the case study network and is connected to the city of Zaandam from which different corridors via Hoorn, Uitgeest and Alkmaar stretch to Enkhuizen (north-east) and Den Helder (north-west). The network consists of 28 stations and is currently served by 8 lines. Each line is operated with a frequency of 1-2 departures per hour but network design is such that the busiest corridors are served by up to 8 trains per hour. This sub-network was selected because it can be assumed to operate relatively independently from the national network yet contains routing alternatives and a network of urban settlements with strong commuting patterns to Amsterdam. Interchanges between the DRT operations within this network and conventional services can be performed at Uitgeest and Amsterdam Central stations.

![Figure 2: Case study network – in relation to the Dutch railway network (left) and zoom-in on the North-Holland subnetwork.](image)

To analyse the rail-DRT capacity requirements for the case study network, we used the existing passenger demand patterns. Passenger demand is obtained from smartcard data records. An important feature of the Dutch smartcard system is that passengers are required to check in and check out for every single journey for fare calculation. An OD matrix is constructed for an average weekday morning rush hour with a total of approximately 7000 passenger trips. Figure 3 visualizes the demand pattern during the morning rush hour before and after the minimum service frequency threshold of one departure per hour per OD pair has been applied. One station, namely Zaandam Kogerveld, is not served at all because the demand to any other station does not amount to a volume that justifies a direct connection as indicated by the absence of any connecting service arcs in Figure 3 (right). A total of 4% of the passenger demand is not served by a direct service in this application.
The same set of parameters as detailed in the Model Implementation and Specification section is used in the base case scenario. Note that the envisaged rail-DRT system is served by vehicles much smaller than current trains (i.e. 24 passengers) and track capacity is significantly higher (i.e. 180 vehicles per hour per single-track) for automated systems. The sensitivity of the results obtained to variations in track capacity – an important design parameter that at the moment cannot be estimated with certainty - is then analysed. Each optimization problem is solved within 30 seconds on a standard PC.

Results and analysis

Base case results

Table 2 presents the cost components and decision variables values for the optimal solution, expressed in hourly terms. Passenger costs are the largest cost component (39%), followed by link (29%) and node (24%) capacity investment cost while operational costs (8%) are considerably lower. Applying the procedure described in the Cost Minimization Problem Formulation section, 191 vehicle trips are required for serving passenger demand.

The values of the link and node capacity decision variables are displayed in Figure 4 as well as their utilization rate – expressed in terms of volume over capacity ratio - of each track segment and station platforms. Most of the stations require either one or two platforms, similar or fewer than their current capacities. Only the busiest stations require three (Alkmaar, Hoorn and Zaandam) or five (Amsterdam Centraal and Amsterdam Sloterdijk) platforms. One might expect the heterogeneity in terms of stopping patterns of the rail DRT system to require a higher capacity. However, the high share of through-going vehicles to dwelling vehicles allows for efficient infrastructure utilization. Utilization rates vary between 0.18 and 0.74. Double track infrastructure is necessary only on the busy corridor between Amsterdam Sloterdijk to Wormerveer (see Figures 2 and 3). This exhibits the important role that network effects play also in the context of DRT services. Single track (per direction) is sufficient for the remainder of the network. Utilization rates vary between 0.01 for the low-demand branch leading to Enkhuizen to 0.63 for the 2X2 tracks between the two Amsterdam stations as well as on the link just after the 2X2 corridor due to the capacity drop to single-track per direction.
Table 2: Case study results expressed in terms of objective function components and related variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Units</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passenger costs</td>
<td>[€1000]</td>
<td>24.32</td>
</tr>
<tr>
<td>Link costs</td>
<td>[€1000]</td>
<td>17.94</td>
</tr>
<tr>
<td>Node costs</td>
<td>[€1000]</td>
<td>14.61</td>
</tr>
<tr>
<td>Operational costs</td>
<td>[€1000]</td>
<td>5.07</td>
</tr>
<tr>
<td>Fleet size</td>
<td>[vehicles]</td>
<td>191</td>
</tr>
<tr>
<td>Offered seat-km</td>
<td>[1000 km]</td>
<td>253.6</td>
</tr>
<tr>
<td>Unserved demand</td>
<td>[-]</td>
<td>4%</td>
</tr>
<tr>
<td>Service effectiveness</td>
<td>[km/hr]</td>
<td>73</td>
</tr>
</tbody>
</table>

Figure 4: Allocated infrastructure capacity in the optimal solution depicted by link thickness and node size and corresponding infrastructure utilization rates shown using the colour scheme

In order to assess the performance of the rail-DRT service, we compare the travel time to those currently offered by the train service obtained from the current timetable. Since free-flow speed was determined based on the timetable travel time between stations, differences in travel time stem from congestion in the DRT network or different routing. The DRT service yields an average travel time of 21 minutes per passenger and saves 230 in-vehicle hours for a single peak hour, an average of 2 minutes per passenger. Travel time increases with the DRT service compared to the current train timetable for several OD-pairs. In all these cases the underlying cause is that while most vehicles are assigned to the shortest path, a minority of the vehicle flow operating between these OD-pairs is routed via the higher capacity route from Zaandam to Hoorn via Alkmaar (western wing of the ring in Figure 4), instead of the
shortest option via Purmerend (eastern wing). The optimization model has found this particular routing beneficial for attaining system optimum at the cost of prolonging travel times for some of the passengers. DRT outperforms the existing service in terms of passenger travel times for short and medium distance trips while no systematic difference is observed for long distance trips.

In addition to travel time savings, most OD-pairs and passengers benefit from a more frequent service with the exception of OD-pairs that are associated with a low-demand station that is currently well-served because of network configuration. Finally, rail-DRT makes the need to transfer and all thus all related passenger costs obsolete.

Sensitivity to changes in track capacity
The optimization model was run with track capacity, $\theta$, of 30, 60, 90, 120, 150 and 180 vehicles per hour. The objective function cost components values – passenger, link capacity and station capacity – are presented in Figure 5. It can be seen that the link costs decrease exponentially with increasing track capacity. Meanwhile, passenger costs appear to decline rapidly first followed by a more modest decrease for $\theta > 90$ vehicles per hour. These results indicate that given the selected passenger value of time, when moving into lower values of $\theta$ the model first tries to limit the increase in travel time at the cost of rising infrastructure costs. Only once the link capacity costs increase sharply, will they be partly traded-off against travel time costs. This is supported by the observation that the average speed is not affected significantly by decreasing track capacity until track capacity becomes smaller than the aforementioned 90 vehicles per hour.

CONCLUSION
This study presents a network cost minimization model for determining the optimal infrastructure capacity and flow distribution of rail-DRT systems. The results of the numerical experiments and the case study application indicate that - unlike conventional rail systems in which stations often are capacity bottlenecks - link capacity properties are more critical for the performance of automated rail-DRT systems than station capacity.

Under the base case parameter settings, the optimal link and station capacity allocation for the case study network is comparable to those currently available in this heavy rail network. Link infrastructure utilization is however higher, in the order of 70% to 85%, compared to approximately 65% with today’s system. The need to invest in link capacity
strongly depends on vehicle characteristics of minimum headway and the speed-density relation. Trading-off link costs against passenger costs is possible in the form of either inducing longer travel time or reducing service frequency by increasing vehicle capacity.

The optimal solution corresponded to the system-optimum flow distribution. Even though model formulation allows for rerouting vehicles to attain global system optimum conditions, only in extreme cases a minority of the vehicle flow is not assigned to the shortest route. The implications of competing DRT service providers can be studied by adjusting the objective function used in this study to guarantee user-equilibrium conditions.

Since this study is pioneering in the public transport system it envisaged, model specification required making a series of assumptions about plausible characteristics of future automated rail-DRT systems. A series of sensitivity analyses was performed to test the impacts of even extreme deviations from the assumed values on model outputs. Future developments are expected to allow finer specifications of technological and service features such as the link speed-density function, the station platforms queuing regime and related assumptions.

This study examined the long-term planning consequences of offering a new transportation technology and service concept. In order to test its potential to substitute the existing rail service, current demand levels were tested in the application. Future research is needed to assess travellers perceptions and preferences towards such services and potential changes in demand patterns that may be caused by the introduction of rail-DRT services. For example, the service might have consequences for station attractiveness due to the increased correlation between service frequency and demand for specific connections. This may also have ramifications for network structure design and service availability, including equity considerations.

ACKNOWLEDGEMENT
The authors are thankful for Dennis Huisman’s suggestions and critical input. Dutch Railways (NS) kindly provided input on operational settings as well as data for the case study.

REFERENCES


