PUBLIC TRANSPORT ON DEMAND

A better match between passenger demand and capacity

M.H. Matena
PUBLIC TRANSPORT ON DEMAND

M.H. (Martijn Hendrik) Matena

Industrial Engineering and Management

Production and Logistic Management

Graduation committee:
Dr. ir. M.R.K. Mes
Dr. ir. J.M.J. Schutten
Dr. ir. H.N. Post

Publication Date:
25 November 2015

University of Twente
Drienerlolaan 5
7522 NB Enschede
The Netherlands
http://www.utwente.edu

Connexxion
Laapersveld 75
1213 VB Hilversum
The Netherlands
http://www.connexxion.nl
INTRODUCTION OF SUMMARY

This project is commissioned by the department Future Technology (FT) at Connexxion. Connexxion is part of the international shareholder Transdev. The goal of Connexxion is to be the best choice in the field of regional passenger and healthcare transport in the Netherlands. Due to confidential reasons only a summary of the non-confidential parts is available, after 27-11-2017 the complete research becomes available for the public. The summary contains in Chapter 1 the problem formulation, Chapter 2 the treated literature, Chapter 3 our solution design, Chapter 4 the performed experiments and the performance evaluation are described. Chapter 5 contains the conclusion and recommendations.

1 PROBLEM FORMULATION

Public transport is becoming more and more important, especially in densely populated cities. The more citizen use public transport, the less congestions and greenhouse gas emissions, which all affect the living environment. But to motivate citizen to use public transport, the public transport service needs to be fast, reliable, flexible, and cheap (Cepeda et al. 2006, Raveau et al. 2011, Schmöcker et al. 2011), while the operators are interested in a profitable system, where wages, and the costs of vehicle usage are low.

The current determination of bus lines are based on average demand. The demand and supply of transport, can have a misfit in the following ways: the offered capacity is too small, so customers cannot be transported to their destination, or the offered capacity is not used, e.g., empty buses. To reduce this misfit a new approach of PT is needed. The new approach must be an on-demand approach, known in literature as Demand Responsive Transport (DRT). A system is categorized as DRT if:

• The service is available to the general public, and is not restricted to a particular group of users.
• The service uses small vehicles like cars or mini vans.
• The route is created between the requested pickup and drop-off location.
• A request is serviced by a vehicle, if a new request is combined in the same vehicle, the route is changed. If the request cannot be combined, the route is not changed.
• The request can be accepted instantly or pre-booked.

Connexxion is developing new methods to improve the service for customers, while increasing the utilization of the vehicles. An ultimate goal is to increase the number of travelers using PT. A new more flexible, on demand driven method could be the solution for this challenge. That is why we formulated the following objective for our research:

Design a solution model that is able to combine and handle real-time DRT requests of customers, with an acceptable service level against minimal costs.
Summary

Connexxion is considering a new solution for the public transport that enables customers to order a ride based on their demand. The current operating bus lines, driving a fixed route, are replaced by a service that creates routes based on the customer demand. A customer has the benefit of no more changes between bus lines, and he or she is able to send a request on a preferred time. We suggest that this model uses predetermined stop locations that are only serviced on-demand. The customer is able to send in a request a short time before the actual start, containing a preferred pickup time. The customer receives a message containing information about the pickup time window, and the latest arrival time.
More than half a century ago, the vehicle routing problem (VRP) was introduced as the truck dispatching problem (Dantzig and Ramser 1959). The VRP is a generic name given to a class of problems involving optimizing vehicle routes. The goal of the VRP is to serve a number of customers with the least amount of vehicles given a set of restrictions, while minimizing the total route costs. We start in Section 2.1 with variation of the VRP. Section 2.2 describes possible methods to solve the VRP. In Section 2.3 the insertion method is discussed. In Section 2.4 describes performance indicators for customer satisfaction. Section 2.5 describes possible implementation issues. In Section 2.6 we formulate our conclusion.

2.1 VARIANTS OF THE VRP

In our research we describe several variants of the VRP:

- A VRP with time windows, meaning a vehicle is only allowed to visit the location between two times (Solomon 1987, Bräysy and Gendreau 2005).
- A Multi-depot VRP, meaning there are more than one depot location, vehicles are allowed to start from multiple location, the basic model assumes a unlimited number of vehicles at each location (Gillett and Johnson 1976).
- A VRP with pickup and delivery, in this VRP type has several categories that set conditions on the pickup and delivery sequences (Nagy and Salhi 2005).
- The Dial-A-Ride-Problem (DARP) is a VRP approach that consists of designing vehicle routes and schedules for a number of customers. The DARP knows two variants: a static and dynamic case. Static means the demand is known beforehand, while the dynamic variant the demand is known a very short time before the actual pickup (Cordeau and Laporte 2003).
- A Demand Responsive Transport model, is a VRP approach quite similar to the DARP, only this variant is more focused on performance for the operator. Like the DARP the DRT has several variants in how pickup and delivery are done (Mulley and Nelson 2009, Wang et al. 2015)
- The Mobility Allowance Shuttle Transit (MAST) is a hybrid transportation system that allows vehicles to deviate from the original route. The original route is servicing the most important stops, while less important stops are only serviced when requested (Quadrifoglio et al. 2006).

2.2 SOLVING A VRP

A VRP problem is a hard combinatorial optimization problem, which is non-deterministic polynomial-time complete (NP-complete) (Nemhauser and Trotter Jr 1975). This means that solving the mathematical problem is hard. Exact algorithms therefore have a slow convergence rate, since a nearly complete enumeration is
Summary


necessary. Solving realistic problem sizes with a constant success rate within an acceptable time is impossible (Cordeau et al. 2002). There are three type of methods discussed in our research:

• Exact methods: A procedure that solves a VRP to optimality, in a limited time period, with a limited problem size, is called exact (Goetschalckx and Jacobs-Blecha 1989). The main disadvantage of an exact approach is the limitation on the problem size, since the problem size grows exponential. In the literature there are several ways to solve the VRP. Some well-known methods are branch-and-bound algorithms, branch-and-cut algorithms, dynamic programming, commodity flow formulations, and set partitioning.

• Classic Heuristics: Classic heuristics are subdivided into two main categories, construction and improvement heuristics. The main difference between these categories is that the construction heuristics, a feasible solution is built by adding routes from “scratch”, whereas the improvement heuristic, also known as local-search, starts with any feasible solution that the heuristic tries to improve. Some well-known methods are cluster-first route-second, the savings algorithm, the set partitioning heuristic (SPH), K-opt, and b-cyclic k-transfer scheme.

• Metaheuristics: All the current formulated metaheuristics allow the exploration of the solution space beyond the local minima. The classic heuristics can give a local minimum as result. Metaheuristics are formulated in such a way that they allow infeasible of inferior solutions in order to escape from local minima. Roughly all metaheuristics can be classified into three categories: local search, population search, and learning mechanisms.

2.3 INSERTION METHOD

Jaw et al. (1986) described, as one of the first, an insertion method on the DARP. They modelled the insertion procedure as follows. Assume \( n \) customers and \( B \) vehicles are given: sort the available requests on the earliest pickup times. Then find all the feasible ways in which a customer can be inserted into the work-schedule of vehicle \( b \in B \). If it is infeasible to insert a customer, then examine the next vehicle \( b + 1 \). If the customer cannot be fitted in the current work-schedules, the algorithm creates a new route that results in minimum additional costs.

2.4 PERFORMANCE INDICATORS FOR CUSTOMER SATISFACTION

To motivate citizens to make use of the PT, instead of their private car, the PT sector needs to satisfy the travellers in some way. Key performance attributes can help in how to measure customer satisfaction. An attribute is a characteristic of an object that can be operationalized. In the related literature, several proposals are made on how to measure the performance of PT. Redman et al. (2013) state that the attributes can roughly be divided into two categories: physical and perceived. The attributes linked to the physical category can be measured without the input of the users of PT, and are made about the impact on the users of PT. The
Summary

perceived attributes are measured using PT users experience. We created Figure 2.1 that defines the most commonly studied PT attributes, where the overall customer satisfaction of all the attributes are represented in the figure.

![Overall Customer Satisfaction Diagram](image)

**Figure 2.1: The service attributes related to PT**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>19.5%</td>
</tr>
<tr>
<td>Frequency</td>
<td>13.6%</td>
</tr>
<tr>
<td>Accessibility</td>
<td>8.2%</td>
</tr>
<tr>
<td>Price</td>
<td>8.3%</td>
</tr>
<tr>
<td>Comfort</td>
<td>12.9%</td>
</tr>
<tr>
<td>Safety</td>
<td>9.8%</td>
</tr>
<tr>
<td>Information provision</td>
<td>11.1%</td>
</tr>
<tr>
<td>Convenience</td>
<td>8.4%</td>
</tr>
<tr>
<td>Reliability</td>
<td>8.1%</td>
</tr>
<tr>
<td>Safety</td>
<td>9.8%</td>
</tr>
</tbody>
</table>

**Figure 2.2: Attribute explanation on customer satisfaction**

Figure 2.2 shows the percentage that explain the influence of the attribute in customer satisfaction. The influences of the attributes are calculated using the highest Pearson correlation if multiple correlation values are known, the percentage is calculated as follows % of attribute = Pearson Correlation²/Total%. Figure 2.2 shows that speed and frequency are the two attributes that have the most influence in the customer satisfaction. From the correlation we see that the pricing of PT is not as important as others. We conclude that customers are willing to accept a higher price, if the attributes with a higher percentage improve (e.g., Speed).
2.5 IMPLEMENTATION ISSUES

DRT is promoted as a PT solution in areas where the traditional services are not economical. Earlier in this chapter, we described the VRP, how to solve the VRP, and what the important satisfaction factors are for passengers. We found the following possible issues: costs and revenue issues, operational issues, institutional issues, economic issues, cultural issues, and information issues.

2.6 CONCLUSION

From the literature we state that our problem can be characterized as a many-to-many DRT or a DARP. For solving the problem we prefer to use a metaheuristic due to the ability to handle a very large number of customers, and to find good solutions in limited time. When the number of locations is small, an exact algorithm can help solving the problem. We use the DRT and DARP, to create a formulation that could handle our problem. Important customer satisfaction attributes are speed, frequency, and comfort. These attributes are important for implementation, to satisfy the customer’s needs. Besides the customer’s needs there are other barriers to overcome, especially the institutional barriers (licenses and insurance), and the economic barriers (subsidy).
3 SOLUTION DESIGN

Inspired by the dial-a-ride problem of Cordeau and Laporte (2003), who formulated a model, that provides a stop-to-stop service with allowed detours to combine rides, we develop a method that is able to handle requests that are only known a short time before the actual pickup. Section 3.1 describes the foundation of our models. In Section 3.2 the models are formulated.

3.1 FOUNDATION

Before we discuss how our model operates, we describe the decisions done by Connexxion. We start with the strategic decisions, followed by the operational decisions.

3.1.1 STRATEGIC DECISIONS

The following decisions on strategic level are made:

- Provide a “when you want and where you want service”, that enables customers to order a ride from and towards a location
- The locations that are served by the system are predetermined.
- A vehicle always starts and ends at the depot location that is located near the centre of the city.
- The fare prices are based on the shortest path or the price is fixed, independent of the ride length.
- The model must perform in a relatively small service area (<80 square KM).

3.1.2 OPERATION DECISIONS

The following decisions on operational level are made, the times are explained by Figure 3.1:

- A requested ride must be know at least $T_R$ minutes (Call time) before the earliest pickup time.
- The customers receive the pickup time window and the latest possible arrival time.
- The customers receive at least $O_t$ minutes (Communication time) a message about the exact pickup time.

![Figure 3.1: Time windows of a ride](image-url)
Figure 3.1 shows a schematic representation of the times of the process. It starts with a request (the blue dot), the request must be done before the call time starts. The call time is the time the model assigns the request to a route. The request contains the earliest pickup time, and the latest pickup time (the dark blue bar). At a certain moment the actual pickup time is communicated (the red square), and the customer enters the vehicle. The time that a customer is in the vehicle is the travel time (the dark green bar). The customer must arrive at the requested location before the latest arrival time, the latest arrival time is determined by the latest pickup time plus the shortest travel time.

For the determination of the maximum detour time, we formulate two models. Model 1 uses a fixed maximum detour time. Independent of the ride length the allowable detour time remains the same. Model 2 uses a detour factor, the allowed detour time is based on the direct ride time times the detour factor. This model results in that all rides have the same relative allowable detour time.

For our models we have made several assumptions:

- Request time: Request must be known before the call time.
- Actual pickup time: The exact pickup time is communicated at least $O_i$ minutes before the actual pickup time, from this moment it is not allowed to insert a location before the communicated location.
- Ride combination: Customers are assigned to a vehicle, when an assignment is not possible the customer is rejected by our models.
- Drive times: The drive times are deterministic.
- Agreements: Customers can only order rides when the service is active, and the customer must be waiting at the pickup location, else the request is not served.
- Arrival time vehicle: The vehicle can only arrive within the time window of the customer, if the vehicle arrives before the earliest pickup moment the vehicle waits until the earliest time of the time window is met.
- Maximum ride time: The maximum ride time is determined beforehand. The maximum ride time is determined in two different ways. 1st the maximum ride time is the shortest path ride time plus a fixed detour time. 2nd the maximum ride time is the shortest path ride time times a factor.
- Boarding Time: The time needed to get in and get off a bus is ignored.
- Route: A route always starts and ends at a depot location, the length of the route can be extended by new requests until either the maximum driving time of the driver is ended, or when there are no more requests assigned.
- Luggage: We assume if a customer carries a regular piece of luggage it causes no capacity problems.
- Time windows violations: When customers receive the time of actual pickup they are assigned a vehicle. In practice it could happen that a vehicle gets delayed, e.g., due to traffic jams. In our
model, real life events that result in delays are not taken into account. So all assigned requests are serviced within the time windows.

3.2 MODEL FORMULATION

In Section 3.2.1 we explain how our formulated models handle the online requests with the use of the insertion method. Section 3.2.2 give a short formulation of model 1 and Section 3.2.3 of model 2.

3.2.1 INSERTION

The idea of the insertion algorithm is to construct the route iteratively by performing an insertion for each request. All infeasible solutions are removed from the solution set. The procedure involves one step for each request (Häme 2011): Perform insertion of the new request to all feasible service sequences with respect to existing customers. The current sequence of locations assigned to a bus cannot be changed.

Objective

The main objective of the insertion method is to minimize the total costs, and reduce the detour time. To achieve this, only feasible insertions are considered. All infeasible insertions are eliminated to reduce the computation time. Narrow time windows result in less computations, so a good solution can be generated fast. Besides reducing the number of computations, another goal is to find a feasible sequence if one exists, and to decide whether or not a request can be accepted by our model.

Formulation

To check whether or not a customer request can be accepted, the insertion checks the feasibility in two phases. The first phase checks if the current sequence, on the precedence, time, and capacity constraints are satisfied. In the second phase, the remaining locations that are not in the sequence are determined by considering the possibilities of adding a remaining location to the sequence.

3.2.2 MODEL 1

Model 1: Here we use the concept of a fixed detour time. The latest pickup time is equal to the earliest pickup time plus flexibility time (TW size). The latest arrival time is equal to the latest possible pickup time plus the direct ride time. Hence the planning flexibility might be considered in two ways, either by waiting at the pickup location or by the detour time.

3.2.3 MODEL 2

Model 2: Here we use the concept of using a factor. The maximum detour is determined by multiplying the direct driving time with the detour factor. All other parameters are determined in the same manner as model 1.
4 PERFORMANCE EVALUATION

To evaluate the performance of our models, we collected data of a relatively small city in the Netherlands. The requests that are used as input for our models, are based on one month OV-chip card data. We are only using requests that travel within the city, all requests using bus lines that leave the city are not used. If all requests are served separately a total distance of 57,069 KM with an average of 2,68KM per request must be travelled. Meaning 2,153:39 operating hours are needed.

Table 4.1 represents the cases for model 1, and Table 4.2 for model 2. In each case only one parameter is adjusted, this parameter is recognized by the arrow and the brackets. The parameter goes from the value before the arrow to the value after the arrow with the step size shown between the brackets. In each table the last row, is a combination of several parameters. For all instances, either a vehicle fleet of nine vehicles with a capacity of three persons (vehicle fleet A), or a vehicle fleet of six vehicles with a capacity of eight persons (vehicle fleet B) is used.

<table>
<thead>
<tr>
<th>Model 1: Fixed Time Window with a fixed detour time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicles</td>
</tr>
<tr>
<td>Case 1</td>
</tr>
<tr>
<td>Case 2</td>
</tr>
<tr>
<td>Case 3</td>
</tr>
<tr>
<td>Case 4</td>
</tr>
<tr>
<td>Case 5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2: Fixed Time Window with a detour factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicles</td>
</tr>
<tr>
<td>Case 6</td>
</tr>
<tr>
<td>Case 7</td>
</tr>
<tr>
<td>Case 8</td>
</tr>
<tr>
<td>Case 9</td>
</tr>
<tr>
<td>Case 10</td>
</tr>
</tbody>
</table>

To measure the performance of each case, we use several measurements. We divide the performance into two categories, performance measures related to the operation (Connexxion), and to customer service. First we treat the performance measures related to the operation, followed by the performance measures related to the customers.

The performance measures for the operation:

- Number of served requests; all the requests that are served by our system.
- Number of not served requests; the requests that could not be served, given the restrictions.
Summary

- Total driven distance; the distance that is driven to serve the requests.
- Total driving time; the time needed to serve all the requests.
- Average occupation; the average number of customers in the vehicle.

The performance measures for the customers:

- Average detour time; the average extra time that is needed to transport customers towards their destination.
- Average waiting time; the average time that a customer needs to wait at the pickup location.

From our experiments we conclude that both models have a different performance, where model 2 perform better in combining requests, model 1 has a better performance for the customers. One of the highest costs are the wages of drivers and the purchase of vehicles. Although the purchase price of the three person vehicle is cheaper, the wage remains the same. In all situations, the performance of vehicle fleet B is better compared to the vehicle fleet A. The less vehicles needed to serve the requests, the higher the revenue becomes while reducing the operating costs, due to the limited number of drivers. To compare both the models, we have summarized the results in Table 4.3. The cells containing a bold font, are the cases with the best performance. Cases 1 to 5 are using model 1, and cases 6-10 are using model 2.

• Table 4.3: Summary of the best settings for each case

<table>
<thead>
<tr>
<th>Case</th>
<th>Vehicle fleet A</th>
<th>Vehicle fleet B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter</td>
<td>Parameter</td>
</tr>
<tr>
<td>1</td>
<td>Call time: 15 minutes</td>
<td>Call time: 15 minutes</td>
</tr>
<tr>
<td>2</td>
<td>Detour time: 6 minutes</td>
<td>Detour time: 7 minutes</td>
</tr>
<tr>
<td>3</td>
<td>TW size: 20 minutes</td>
<td>TW size: 20 minutes</td>
</tr>
<tr>
<td>4</td>
<td>11 vehicles</td>
<td>6 vehicles</td>
</tr>
<tr>
<td>5</td>
<td>Call time: 10 minutes; Detour time: 6 minutes</td>
<td>Call time: 20 minutes; Detour time: 9 minutes</td>
</tr>
<tr>
<td>6</td>
<td>Call time: 15 minutes</td>
<td>Call time: 15 minutes</td>
</tr>
<tr>
<td>7</td>
<td>Detour factor: 0.8</td>
<td>Detour factor: 1.2</td>
</tr>
<tr>
<td>8</td>
<td>TW size: 20 minutes</td>
<td>TW size: 20 minutes</td>
</tr>
<tr>
<td>9</td>
<td>10 vehicles</td>
<td>6 vehicles</td>
</tr>
<tr>
<td>10</td>
<td>Call time: 10 minutes; Detour factor: 0.8</td>
<td>Call time 20 minutes; detour factor: 1.2</td>
</tr>
</tbody>
</table>

Cases 1 and 6 are showing the best performance using a call time of 15 minutes. A 15 minutes call time, is enough to service the most requests, a longer call time result in assigning request in sequences with less performance.

Case 2 shows a small increase in the maximum detour time, results in a better performance for both the vehicle fleets. When we vehicle fleet B is used the detour time one minute higher. This is due to the
capacity of the vehicles, the larger vehicles are able to combine more requests, and combining requests require planning flexibility. Case 7 shows a similar higher detour factor for vehicle fleet B.

Cases 3 and 7 are showing the best balanced performance using a pickup TW of 20 minutes, independent of the vehicle fleet. More planning flexibility result in a better performance for Connexxion, but reduces the performance for the customers.

Cases 4 and 7 show a large difference in the number of vehicles to serve the same number of requests. An acceptable performance for vehicles with a capacity of eight persons is achieved, using six vehicles, while at least ten vehicles with a capacity of three persons are needed to achieve an acceptable result. Using less vehicles drastically reduces the operating costs, but we must keep in mind that we need to serve at least 95% of all request.

Case 5 shows that the best performance is found with a call time of 10 minutes and a detour time of 6 minutes using vehicle fleet A. Using vehicle fleet B a call time of 20 minutes and a detour time of 9 minutes are showing the best performance. Vehicle fleet A shows a smaller call and detour time. We state this is due the higher number of available vehicle, increasing the planning flexibility. The same reason holds for case 10.

Table 4.3 shows, that in some cases the choice for a model varies between the cases. When we use vehicle fleet A, we see that in 3 cases model 2 has the better performance. When we use vehicle fleet B, we see the opposite result. We think the best overall performance is found using vehicles with a capacity of eight persons. Due to the fact that in all cases the eight person vehicles, are outperforming the three person vehicles on all the performance indicators for the operation. The performance for the customers is only slightly reduced.
5 CONCLUSION AND RECOMMENDATIONS

In this chapter, we mention the conclusions of our research in Section 5.1 and our recommendations in Section 5.2.

5.1 CONCLUSION

In contrast to prevailing methods in literature, we use a small service area, in which all the stops can be served from one depot location without violating any restriction. In the small service area, all customers that request a ride within the area are serviced on-demand. When a customer wants to leave the service area he or she can be transported towards a public transport hub (e.g., a train station). Another difference compared to the literature is the way the latest arrival time is determined. In our model the latest arrival time is determined by using the latest possible pickup time plus the shortest travel time. This method of determining the latest arrival time reduces the system flexibility but, it improves the speed of transport for the customers.

Our experiments show that changing parameters have a significant influence on the performances of our models. Our results show that only 9 small vehicles with a capacity of eight persons are needed, to serve all requests using one of our models. When we use model 2 it is possible to serve all the request driving 36,714 km with a corresponding 1,464 hours in a month. The results show us the effects when serving the customers on demand. Based on the results we believe that the use of on-demand transport is possible, and profitable for Connexxion. Although our model is not extensively tested, we believe if makes a valuable contribution in getting more insight of the possibilities of on-demand transport.

5.2 RECOMMENDATIONS AND FURTHER RESEARCH

For further research, we suggest to find out the impact of using flexible vehicle locations instead of using one depot. We also suggest study the possible combination of on-demand transport and bus lines. Another suggestion is to take online events into account (e.g., accidents, traffic jams, and rush hours) and the use of stochastic travel times. We suggest that a possible improvement of our model is achieved by reassigning the customers that did not already received an actual pickup time. To increase the possible service level for the customers, we suggest a case that allows the customers to change the parameters. We state that customers that cannot be served in within the given restrictions, are rejected, it might be helpful to see the effect of offering the rejected request an alternative pickup time.
REFERENCES


Summary


