



Doctoral Thesis

## Robust Routing in Urban Public Transportation Networks

**Author(s):**

Pröger, Tobias

**Publication Date:**

2016

**Permanent Link:**

<https://doi.org/10.3929/ethz-a-010797721> →

**Rights / License:**

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

DISS. ETH No. 23513

# Robust Routing in Urban Public Transportation Networks

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZÜRICH  
(Dr. sc. ETH Zurich)

presented by  
TOBIAS PRÖGER  
Dipl.-Inform., Technische Universität Dortmund  
born on September 5, 1987  
citizen of Germany

accepted on the recommendation of  
Prof. Dr. Peter Widmayer, examiner  
Prof. Dr. Joachim M. Buhmann, co-examiner  
Prof. Dr. Anita Schöbel, co-examiner

2016

# Abstract

In public transportation, it is a well-studied problem to compute journeys that leave a given departure stop  $d$  at a given departure time  $t_D$ , and that are planned to reach a given target stop  $t$  as early as possible. A related problem asks, given a departure stop  $d$ , a target stop  $t$  and a latest allowed arrival time  $t_A$ , to compute a journey that is planned to leave  $d$  as late as possible and reaches  $t$  on time (i.e., not later than  $t_A$ ). If the real arrival and departure times of each vehicle correspond to the planned ones then it is certainly sufficient to use the planned timetable for computing such journeys.

In reality, however, delays in public transportation are omnipresent due to traffic congestions, road or track work, passengers boarding and leaving vehicles, etc. Hence, the real vehicle movements often deviate substantially from the scheduled ones. This is problematic because not only does the travel time of single vehicles increase, but different arrival times may also lead to failed transfers between vehicles of different lines, increasing further the overall travel time. The first part of this thesis therefore studies the problem of finding journeys that are *robust* against delays. More concretely, we assume that a departure stop  $d$ , a target stop  $t$  and a latest allowed arrival time  $t_A$  are given, and our goal is compute journeys that likely reach  $t$  on time.

Instead of modeling delays explicitly by artificial probability distributions, we will use historic delay data and look for journeys that performed well w.r.t. the given past data. For this purpose we develop a solution concept that separates the topological structure (i.e., the lines and the transfers used) from concrete travel times, and propose a three-stage approach for finding robust journeys. As a first step, we generate all *dt-routes* (i.e., feasible sequences of lines) whose number of transfers do not exceed a given threshold. After that, we compute one or more robust route(s) and corresponding departure times. Finally, appropriate transfers are computed using the planned timetable.

We will develop various methods for computing robust routes. Simple methods ensure robustness only by adding additional buffer time at each transfer. A more sophisticated method, originally proposed by Buhmann et al. (2013) for combinatorial optimization, will be adapted to public transportation. Since this method does only take historic data of two different days into account where in reality often more data is available, we study possible generalizations of this method to incorporate data of multiple past days. We will also adapt a mean-risk model originally proposed by Lim et al. (2012) for private transportation to public transportation. This method chooses routes that minimize a weighted sum of the mean and the standard deviation of the minimal times

one had to leave in advance, where times are measured w.r.t. the data of the past days. We will then experimentally evaluate these methods on real-world data of the public transportation network of Zürich.

In the second part of the thesis, we will further investigate the aforementioned method by Buhmann et al. (2013) for robust optimization. In this method it is assumed that a problem generator generates similar problem instances that differ slightly due to noise. Assuming that nothing is known about the type of the noise or the problem generator itself except for exactly two instances generated by it, the goal is to identify solutions that are likely to perform well for future, yet unknown instances generated by the same generator. For this purpose, Buhmann et al. (2013) introduced the concept of a  $\gamma$ -approximation set which contains all  $\gamma$ -optimal solutions. They then compute the  $\gamma$  that maximizes the *similarity* of the given two instances, which relates the actual size of the intersection of the  $\gamma$ -approximation sets to the expected number of solutions in the intersection. In this thesis we will propose and experimentally evaluate models in which this method outperforms other methods. As a first step towards a formal investigation of the method, we will give an analytic expression to estimate the similarity of two instances, and experimentally evaluate the quality of the estimation. Furthermore, we will discuss possible generalizations to incorporate multiple input instances.

# Zusammenfassung

Ein oft untersuchtes Problem im öffentlichen Verkehr ist die Berechnung von Reisen, die einen gegebenen Abfahrtsort  $d$  zu einer gegebenen Abfahrtszeit  $t_D$  verlassen, und die einen gegebenen Zielort  $t$  geplant so früh wie möglich erreichen. In einem verwandten Problem soll für einen gegebenen Abfahrtsort  $d$ , einen gegebenen Zielort  $t$  und eine Zielankunftszeit  $t_A$  eine Reise berechnet werden, die  $d$  so spät wie möglich verlässt und  $t$  dennoch pünktlich (d.h., nicht später als  $t_A$ ) erreicht. Stimmen die realen Ankunfts- und Abfahrtszeiten jedes Fahrzeugs mit den geplanten überein, dann ist es sicherlich ausreichend, den Fahrplan als Berechnungsgrundlage für solche Reisen zu benutzen.

In der Realität sind Verspätungen allerdings aufgrund von Verkehrsüberlastungen, Strassen- oder Gleisbauarbeiten, ein- oder aussteigenden Passagieren, usw., allgegenwärtig. Daher weichen die realen Bewegungen einzelner Fahrzeuge von den geplanten deutlich ab. Dies ist problematisch, da sich nicht nur die Reisezeit einzelner Fahrzeuge erhöht, sondern geänderte Ankunftszeiten auch zu verpassten Anschlüssen zu Fahrzeugen anderer Linien führt, was die Gesamtreisezeit weiter erhöht. Der erste Teil der vorliegenden Arbeit untersucht daher das Problem der Berechnung von Reisen, die *robust* gegenüber Verspätungen sind. Konkret nehmen wir an, dass ein Abfahrtsort  $d$ , ein Zielort  $t$  und eine Zielankunftszeit  $t_A$  gegeben sind, und unsere Aufgabe besteht in der Berechnung von Reisen die  $t$  wahrscheinlich pünktlich erreichen.

Anstatt Verspätungen explizit durch künstliche Wahrscheinlichkeitsverteilungen zu modellieren, werden wir historische Verspätungsdaten benutzen, und Reisen suchen die gute Eigenschaften in Bezug auf die gegebenen Daten aufweisen. Zu diesem Zweck entwickeln wir ein Lösungskonzept das die topologische Struktur (d.h., benutzte Linien und Umstiege) von konkreten Reisezeiten trennt, und schlagen einen dreistufigen Ansatz zur Berechnung robuster Reisen vor. In einem ersten Schritt generieren wir alle *dt-Routen* (d.h., zulässige Liniensequenzen), deren Anzahl von Umstiegen einen vorgegebenen Grenzwert nicht übersteigt. Danach berechnen wir eine oder mehrere Route(n) und entsprechende Abfahrtszeiten. Abschliessend benutzen den Fahrplan zur Berechnung geeigneter Umstiege.

Wir werden verschiedene Methoden zur Berechnung robuster Routen entwickeln. Einfache Methoden stellen Robustheit lediglich durch die Benutzung zusätzlicher Pufferzeit bei jedem Umstieg sicher. Eine höher entwickelte Methode, ursprünglich von Buhmann et al. (2013) für kombinatorische Optimierung entwickelt, wird in das Szenario der Berechnung robuster Routen übertragen. Während in der Realität oftmals Daten vieler vergangener Tage vorhanden

sind, kann die genannte Methode lediglich die Daten zweier vergangener Tage einbeziehen. Wir werden daher mögliche Verallgemeinerungen der Methode untersuchen, sodass die Daten mehrerer vergangener Tage einbezogen werden können. Weiterhin werden wir ein Risikomodell, ursprünglich von Lim et al. (2012) für den privaten Verkehr entwickelt, in unser Szenario übertragen. Diese Methode wählt Routen aus, die eine gewichtete Summe des Mittelwerts und der Standardabweichung der minimalen Zeiten, um die man in Bezug auf die Daten der vergangenen Tage im Voraus hätte abfahren müssen, minimieren. Wir werden die Methoden dann experimentell auf echten Verspätungsdaten des ÖPNV-Netzes von Zürich vergleichen.

Im zweiten Teil der Arbeit werden wir die vorhergenannte Methode von Buhmann et al. (2013) zur robusten Optimierung weiter untersuchen. In dieser Methode wird angenommen, dass ein Problemgenerator ähnliche Probleminstanzen generiert, die sich aufgrund von Rauschen ein wenig unterscheiden. Unter der Annahme, dass weder die Art des Rauschens noch der Problemgenerator selbst bekannt sind und lediglich zwei vom Generator erzeugte Instanzen zur Verfügung stehen, besteht die Aufgabe nun in der Identifizierung von Lösungen die wahrscheinlich für weitere, derzeit unbekannte, vom Generator erzeugte Instanzen gute Eigenschaften aufweisen. Zu diesem Zweck führten Buhmann et al. (2013) das Konzept einer  $\gamma$ -Approximationsmenge ein, die alle  $\gamma$ -optimalen Lösungen enthält. Sie berechnen dann den Wert  $\gamma$ , der die *Ähnlichkeit* der gegebenen zwei Instanzen maximiert. Dieses Mass setzt die tatsächliche Grösse des Schnitts der  $\gamma$ -Approximationsmengen in Bezug zur erwarteten Anzahl der Lösungen im Schnitt. In dieser Arbeit werden wir Modelle vorschlagen, in denen diese Methode bessere Ergebnisse als andere Methoden erzielt, und diese Modelle experimentell evaluieren. Im Hinblick auf eine formale Untersuchung der Methode werden wir einen analytischen Ausdruck zur Schätzung der Ähnlichkeit zweier Instanzen angeben, und die Qualität der Schätzung experimentell evaluieren. Weiterhin werden wir mögliche Verallgemeinerungen der Methode diskutieren, damit mehrere Eingabeinstanzen berücksichtigt werden können.