Konrad-Zuse-Zentrum für Informationstechnik Berlin
Takustr. 7, D-14195 Berlin-Dahlem

Martin Grötschel Andreas Löbel Manfred Völker

Optimierung des Fahrzeugumlaufs im Öffentlichen Nahverkehr

Preprint SC 96–8 (August 1996)

Optimierung des Fahrzeugumlaufs im Öffentlichen Nahverkehr

M. Grötschel[1], *A. Löbel*[1] *und M. Völker*[2]

[1] Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Takustraße 7, D-14195 Berlin-Dahlem, E-mail: `[name]@zib.de`, URL: `http://www.zib.de/`
[2] HanseCom GmbH, Spohrstraße 6, 22083 Hamburg

Abstract. This paper addresses the problem of scheduling vehicles in a public mass transportation system. We show how this problem can be modelled as a special multicommodity flow problem and outline the solution methodology we have developed. Based on polyhedral investigations, we have designed and implemented a branch&cut algorithm and various heuristics with which real vehicle scheduling problems of truely large scale can be solved to optimality. We describe some implementation issues and report computational results.

Diese Veröffentlichung wurde durch das BMBF[a]-Projekt "Anwendungsorientierte Verbundvorhaben auf dem Gebiet der Mathematik" im Themenbereich "Mathematische Optimierung und Steuerung technischer Systeme" gefördert.

[a] Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie

1 Einleitung

Der Betrieb eines öffentlichen Nahverkehrssystems kann – aus der abstrakten Sicht eines Mathematikers – als ein gigantisches Optimierungsproblem mit komplexen Nebenbedingungen aufgefaßt werden. Wir diskutieren in diesem Aufsatz eine wichtige Komponente, die Fahrzeugumlaufplanung, und stellen das Problem anhand der Busumlaufplanung dar.

Wir gehen davon aus, daß ein Fahrplan erstellt wurde. Zur Durchführung aller Fahrten des Fahrplanes müssen Busse bereitgestellt werden, wobei berücksichtigt werden muß, daß gewisse Busse (Doppeldecker, Gelenkbusse) nicht alle Linien bedienen können. Das Ziel ist, so wenige Fahrzeuge wie möglich einzusetzen und die Kosten für "Leerfahrten" so gering wie möglich zu halten. (Dabei wird eine Dienst- und Dienstreihenfolgeplanung noch nicht berücksichtigt. Sie geschieht in einem nachgeordneten Optimierungsschritt.)

Zur Lösung derartiger Umlaufplanungsprobleme werden vielfach noch manuelle, z. T. computergestützte Planungsmethoden verwendet. Fortgeschrittene

Verkehrsbetriebe benutzen heuristische Algorithmen, die auf mathematischen Modellen basieren. Durch enorme Fortschritte in der mathematischen Methodik und der Computertechnik ist es heute möglich geworden, die riesigen Optimierungsprobleme, die in diesem Bereich entstehen, exakt zu lösen.

Löbel und Strubbe [LS] berichten, wie unser Lösungsansatz bei den Berliner Verkehrsbetrieben (BVG) für die praktische Nutzung umgesetzt wird. Bereits im Einsatz befindliche, auf Heuristiken basierende Systeme (wie z. B. HOT II) werden von Petzold und Schütze [PS] und Schütze und Völker [SV] beschrieben. HOT II ist ein Softwarepaket der HanseCom GmbH, die eine Tochter der Hamburger Hochbahn AG und von SNI ist.

2 Das Fahrzeugumlaufplanungsproblem

Wir wollen nun einige Begriffe definieren, die uns eine korrekte mathematische Formulierung des Umlaufplanungsproblems erlauben, aber auch gleichzeitig die Behandlung von Sonderfällen und Nebenbedingungen ermöglichen.

Ein *Depot* ist eine Menge von Fahrzeugen, die im Rahmen der vorzunehmenden Planung als gleichwertig angesehen werden. Zu jedem Depot gehört ein Startpunkt und ein Endpunkt, wo jedes eingesetzte Fahrzeug seinen Einsatz beginnt bzw. beendet. Es ist Aufgabe des betrieblichen Planers, eine für seinen Verkehrsbetrieb angemessene Einteilung des Fahrzeugbestandes in Depots vorzunehmen. In Extremfällen können alle Fahrzeuge zu einem Depot zusammengefaßt oder jedes Fahrzeug für sich allein als Depot aufgefaßt werden. In der Regel wird ein Depot aus allen Fahrzeugen eines Typs (oder aus Fahrzeugen, die für die vorliegende Planung als homogen angesehen werden) bestehen, die einem Betriebshof zugehören. In diesem Fall sind Start- und Endpunkt ein und derselbe Betriebshof.

Aus dem Fahrplan (und dem zugrundeliegenden Linienplan) werden die *Fahrgastfahrten* (inkl. Sonderfahrten) mit ihren Anfangs- und Endhaltepunkten und Start- und Endzeiten bestimmt. Jeder Fahrgastfahrt wird eine *Depotgruppe* zugeordnet, die aus allen Depots besteht, die diese Fahrgastfahrt durchführen können. Depotspezifische Kosten zur Durchführung dieser Fahrt können ebenfalls angegeben werden. Die *Hauptaufgabe der Umlaufplanung* besteht nun darin, für jede Fahrgastfahrt des Fahrplans ein Fahrzeug bereitzustellen, das zu ihrer Depotgruppe gehört. Die dazu benötigte Fahrzeugflotte muß alle Nebenbedingungen erfüllen, und die operativen Kosten sollen "so gering wie möglich" sein. Hierbei können mehrere Zielsetzungen verfolgt werden. Manche Verkehrsbetriebe wollen den vorhandenen Fahrzeugpark so einsetzen, daß die operativen Kosten minimiert werden, andere bestimmen zunächst die minimale Zahl von Fahrzeugen, mit denen der Fahrplan bedient werden kann, und danach einen kostenminimalen Umlaufplan mit der geringsten Fahrzeuganzahl.

Um diese Zielsetzungen mathematisch definieren zu können, führen wir weitere Begriffe ein. Neben den zu bedienenden Fahrgastfahrten betrachten wir weitere Typen von Fahrten: *Einsetzfahrten* führen vom Startpunkt eines Depots zum Anfangshaltepunkt einer Fahrgastfahrt, *Aussetzfahrten* führen vom Endhaltepunkt einer Fahrgastfahrt zum Endpunkt eines Depots, *Kopplungen* führen vom Endhaltepunkt einer Fahrgastfahrt zum Anfangshaltepunkt einer anderen Fahrgastfahrt. Zur Bezeichnungsvereinfachung nennen wir eine Fahrt, die keine Fahrgastfahrt ist, *Leerfahrt*.

In unserem mathematischen Modell ist es möglich, für jede Fahrgast- und Leerfahrt depotspezifische *Gewichte* anzugeben, z. B. die strecken- und/oder zeitabhängigen Kosten zur Bedienung der Fahrt mit einem bestimmten Fahrzeugtyp. In die Gewichtung einer Leerfahrt können die Kosten für die Standzeit mit oder ohne Fahrer und die Kosten für eine erforderliche Verbindungsfahrt eingehen. Genaue Kostendefinitionen müssen betriebsspezifisch erfolgen, wobei die Zielsetzungen des Betriebes quantitativ umzusetzen sind.

Ein *Umlauf* ist eine Kette von Fahrten beginnend mit einer Einsetzfahrt, auf die alternierend Fahrgastfahrten und Kopplungen folgen und die mit einer Aussetzfahrt endet. Ein Umlauf führt also immer vom Anfangspunkt eines Depots zum Endpunkt desselben Depots und wird von genau einem Fahrzeug durchgeführt. Das *Gewicht eines Umlaufs* ist die Summe der Gewichte der Fahrgast- und Leerfahrten des Umlaufs.

Hauptziel der gängigen Methoden zur Umlaufoptimierung ist die Minimierung der Anzahl der Umläufe. In allen uns bekannten Veröffentlichungen zu diesem Thema, welche über Anwendungen im Echteinsatz berichten (siehe z. B. [DM] oder [DMS]), werden allerdings nur Kopplungen betrachtet, die eine vorgegebene maximale Wendezeit nicht überschreiten. Gleichzeitig wird behauptet, daß durch die Bestimmung der minimalen Anzahl von Umläufen auch die minimale Anzahl der benötigten Fahrzeuge bestimmt ist. Dies ist jedoch i. a. falsch, siehe [Lo1].

Um den minimalen Fahrzeugbedarf bestimmen zu können, führen wir immer dann zusätzliche Kopplungen ein, wenn es möglich ist, zwei Fahrgastfahrten über eine Aussetzfahrt gefolgt von einer Einsetzfahrt mit ein und demselben Fahrzeug zu bedienen und wenn eine Mindeststandzeit im Depot (oder auf einem Abstellplatz) eingehalten wird. Diese zusätzlichen Kopplungen nennen wir *Betriebshof-Fahrten*. Einer Betriebshof-Fahrt wird eine den betrieblichen Zwecken entsprechende “Gewichtung” zugewiesen. Die Einbeziehung der Betriebshof-Fahrten führt zu einer enormen Erhöhung der Problemgrößen. Uns ist jedoch keine andere (einigermaßen vernünftige) Möglichkeit bekannt, die Minimierung der Fahrzeuganzahl mathematisch exakt zu modellieren.

Wir könnten jetzt den oben geprägten Begriff Umlauf durch Einbeziehung der Betriebshof-Fahrten erweitern. Dies ist jedoch bei Verkehrsbetrieben nicht üblich. Daher führen wir einen neuen Begriff ein. Eine Folge von nacheinander durchgeführten Umläufen nennen wir eine *Umlaufkette*. Damit ist die

Minimierung der Fahrzeuganzahl äquivalent zur Minimierung der Anzahl der Umlaufketten. Das *Gewicht einer Umlaufkette* ist die Summe der Gewichte seiner Fahrgast- und Leerfahrten.

Mit der oben vorgenommenen Begriffsbildung können wir das zunächst nur verbal formulierte *Ziel der Umlaufplanung* quantifizieren: *Es ist eine gewichtsminimale Menge von Umlaufketten zu bestimmen, so daß jede Fahrgastfahrt in genau einer Umlaufkette enthalten ist.*

Auch wenn die Problemstellung der Umlaufplanung bereits recht kompliziert erscheint, so sind doch noch eine Reihe von Sonderfällen zu beachten, die bei manchen Verkehrsbetrieben auftreten bzw. deren Berücksichtigung gewünscht wird. Wir listen hier nur Stichworte auf, ohne die zugehörigen (umfangreichen) Details zu erläutern:

Vermeidung von **Linienwechsel** (für Fahrzeuge, bei denen ein Linienwechsel aufwendige Umrüstungen erfordert); Vermeidung von **Fahrtartenwechsel** (Stamm- und Ergänzungsfahrten wie Schulbusfahrten sollen nicht gemischt werden); Berücksichtigung einer **maximale Wendezeit**; Festlegung von **Erst- und Letztfahrten**; Berücksichtigung von **Bereitschaften** und **vordefinierten (Teil-)Umläufen**; geplante **Depotwechsel**.

Alle diese Anwenderanforderungen können in unserem Modell durch Streichung bestimmter Kopplungen, geeignete Einführung neuer Kopplungen bzw. Modifikation von Gewichten berücksichtigt werden.

Folgende Anforderungen können wir in unserem Modell nicht berücksichtigen: **Maximale Umlaufdauer** und/oder **Umlauflänge** für bestimmte Fahrzeugtypen und **maximale Anzahl von Linienwechseln pro Umlauf**. Wir können jedoch durch heuristische (siehe z. B. [FP]) oder manuelle Nachbearbeitung (wie das i. a. üblich ist) weiche oder harte Nebenbedingungen dieser Art einbeziehen.

3 Das Mathematische Modell

Mit der in Abschnitt 2 eingeführten Begriffsbildung formulieren wir nun das mathematische Modell des Umlaufplanungsproblems. Wir beginnen mit einer graphentheoretischen Beschreibung. Hierzu führen wir einen Digraphen $D := (V, A)$ ein, dessen Knoten V und Bögen A wie folgt definiert sind:

Mit $\mathcal{D}$ bezeichnen wir die Menge der Depots. Für jedes Depot $d \in \mathcal{D}$ bezeichnen wir mit d^+ seinen Startpunkt und mit d^- seinen Endpunkt; wir setzten $\mathcal{D}^- := \{d^- \mid d \in \mathcal{D}\}$ und $\mathcal{D}^+ := \{d^+ \mid d \in \mathcal{D}\}$. Die Menge der Fahrgastfahrten bezeichnen wir mit $\mathcal{T}$. Für jede Fahrgastfahrt $t \in \mathcal{T}$ führen wir für den Anfangshaltepunkt einen Knoten t^- und für den Endhaltepunkt einen Knoten t^+ ein und definieren $\mathcal{T}^- := \{t^- \mid t \in \mathcal{T}\}$ und $\mathcal{T}^+ := \{t^+ \mid t \in \mathcal{T}\}$. Die Knotenmenge V des Digraphen D besteht aus den Start- und Endpunkten der Depots und allen Anfangs- und Endhaltepunkten der Fahrgastfahrten, d. h.

$$V := \mathcal{D}^+ \cup \mathcal{D}^- \cup \mathcal{T}^+ \cup \mathcal{T}^-.$$

Für jede Fahrgastfahrt $t \in \mathcal{T}$ bezeichnen wir mit $G(t) \subseteq \mathcal{D}$ die *Depotgruppe* von t, d. h. die (nichtleere) Menge aller Depots, deren Fahrzeuge die Fahrt t durchführen können. Die Menge aller Fahrgastfahrten, die vom Depot $d \in \mathcal{D}$ bedient werden können, bezeichnen wir mit $\mathcal{T}_d$, d. h. $\mathcal{T}_d := \{t \in \mathcal{T} \mid d \in G(t)\}$; analog definieren wir $\mathcal{T}_d^- := \{t^- \mid t \in \mathcal{T}_d\}$, $\mathcal{T}_d^+ := \{t^- \mid t \in \mathcal{T}_d\}$ und $V_d := \{d^+, d^-\} \cup \mathcal{T}_d^- \cup \mathcal{T}_d^+$.

Wir kommen zu den Bögen. Für jedes Depot $d \in \mathcal{D}$ führen wir einen *Rückführungsbogen* (d^-, d^+) ein, mit dessen Hilfe wir Depotkapazitäten kontrollieren können. Pro Depot $d \in \mathcal{D}$ führen wir für jede Fahrgastfahrt, Einsetzfahrt, Aussetzfahrt und Kopplung jeweils einen Bogen ein, was wie folgt geschieht:

$$\begin{aligned}
A_d^{\text{gast}} &:= \{(t^-, t^+) \mid t \in \mathcal{T}_d\} \text{ (Fahrgastfahrten)},\\
A_d^{\text{ein}} &:= \{(d^+, t^-) \mid t^- \in \mathcal{T}_d^-\} \text{ (Einsetzfahrten)},\\
A_d^{\text{aus}} &:= \{(t^+, d^-) \mid t^+ \in \mathcal{T}_d^+\} \text{ (Aussetzfahrten)},\\
A_d^{\text{kopp}} &:= \dot{\bigcup}_{\text{Kopplungen } (p,q) \text{ für } d} \{(p^+, q^-)\},\\
A_d^{\text{leer}} &:= A_d^{\text{ein}} \dot{\cup} A_d^{\text{aus}} \dot{\cup} A_d^{\text{kopp}} \text{ (Leerfahrten)}.
\end{aligned}$$

Wir erläutern den Gebrauch der Bezeichnungen. Wenn wir von einem Bogen (t^-, t^+) sprechen, so müßten wir eigentlich noch vermerken, bezüglich welchen Depots d der Depotgruppe $G(t)$ der Bogen benutzt wird. Dies sollte i. a. implizit klar sein, zusätzliche Indizes würden die Notation noch häßlicher machen. Wenn wir zwei Bogenmengen A_d^{gast} und $A_{d'}^{\text{gast}}$ vereinigen, so entstehen dabei für jede Fahrgastfahrt $t \in \mathcal{T}_d \cap \mathcal{T}_{d'}$ zwei parallele Bögen. Wir operieren hier also mit "disjunkten Vereinigungen". Analoges gilt für die Kopplungen. Vom Endhaltepunkt p^+ einer Fahrgastfahrt p kann es zum Anfangshaltepunkt q^- einer anderen Fahrgastfahrt q auch mehrere Kopplungen geben. Dies kann auch bereits bezüglich eines Depots der Fall sein, jedoch werden sich dann i. a. die Gewichte der parallelen Kopplungen unterscheiden. Auch hier ersetzen wir nicht zwei parallele Bögen durch einen Bogen, wir vereinigen die Bogenmengen disjunkt.

Für jedes Depot $d \in \mathcal{D}$ erhalten wir damit die folgende Menge von Bögen: $A_d := A_d^{\text{gast}} \dot{\cup} A_d^{\text{leer}} \dot{\cup} \{(d^-, d^+)\}$. Die Bogenmenge A des Digraphen für das Umlaufproblem ist dann die disjunkte Vereinigung all dieser Bogenmengen:

$$A := \dot{\bigcup}_{d \in \mathcal{D}} A_d.$$

Für jeden Bogen aus A führen wir nun noch Bezeichnungen für seine Kosten und Kapazitätsschranken ein. Für den Bogen $a \in A_d$ bezeichnen wir mit $c_a^d \in \mathbb{Q}$ seine Kosten und mit l_a^d bzw. u_a^d seine untere bzw. obere Kapazitätsschranke. Für einen Rückführungsbogen $a = (d^-, d^+)$ ist l_a^d in der Regel die untere Kapazitätsschranke des Depots d und u_a^d die maximale Kapazität von d. Für die übrigen Bögen hat l_a^d den Wert null und u_a^d den Wert eins.

Analog führen wir für jedes Depot d und jeden Bogen $a \in A_d$ eine ganzzahlige Variable x_a^d ein; x_a^d muß die vorgegebenen Kapazitätsschranken erfüllen:

$$l_a^d \quad \leqslant \quad x_a^d \quad \leqslant \quad u_a^d. \tag{1}$$

Aufgrund unserer Annahmen ist eine Variable x_a^d eine Entscheidungsvariable, die angibt, ob ein Fahrzeug des Depots d die Fahrt a durchführt, außer a ist der Rückführungsbogen des Depots d; in diesem Fall zählt x_a^d die eingesetzten Fahrzeuge des Depots d.

Damit ein ganzzahliger Vektor $x := \left((x_a^d)_{a\in A}\right)_{d\in\mathcal{D}} \in \mathbb{R}^A$ eine zulässige Lösung des Umlaufplanungsproblems darstellt, muß er neben den Kapazitätsschranken (1) offensichtlich noch die folgenden Bedingungen erfüllen:

$$\sum_{a\in\delta^-(v)\cap A_d} x_a^d \quad - \sum_{a\in\delta^+(v)\cap A_d} x_a^d \quad = \quad 0 \quad \forall\, v\in V \;\; \forall\, d\in\mathcal{D}, \tag{2}$$

$$\sum_{d\in G(t)} x_{(t^-,t^+)}^d \quad = \quad 1 \quad \forall\, t\in\mathcal{T}, \tag{3}$$

wobei $\delta^-(v) := \{(i,j)\in A \mid j=v\}$ und $\delta^+(v) := \{(i,j)\in A \mid i=v\}$. Die Gleichungen (2) besagen (in der Sprache der Fahrzeugeinsatzplanung), daß alle Fahrzeuge eines Depots d, die einen Knoten $v\in V$ erreichen, diesen Knoten auch wieder verlassen müssen. Die Gleichungen (3) erzwingen, daß jede Fahrgastfahrt t genau einmal durchgeführt wird, wobei natürlich nur Fahrzeuge der zugehörigen Depotgruppe zugelassen sind.

Damit können wir nun das Umlaufplanungsproblem als das folgende lineare ganzzahlige Optimierungsproblem darstellen:

$$\min_{\substack{x\,\in\,\mathbb{Z}^A:\\ x \text{ erfüllt } (1\text{–}3)}} \sum_{d\in\mathcal{D}} \sum_{a\in A_d} c_a^d x_a^d. \tag{4}$$

Dieses ganzzahlige Programm (4) ist ein spezielles Mehrgüterflußproblem. Betrachten wir jedes Depot d für sich, so beschreibt ein Lösungsvektor x^d wie sich Fahrzeuge vom Startpunkt d^+ durch den Digraphen (V_d, A_d) verteilen, dann im Endpunkt d^- zusammenströmen und über den Rückführungsbogen (d^-, d^+) wieder dem Startpunkt zugeführt werden. In der Sprache der Netzwerkflußtheorie wird x^d ein *Fluß* oder genauer eine *Zirkulation* (des "Gutes d") im Digraphen (V_d, A_d) genannt. Die Schwierigkeit besteht darin, daß die Zirkulationen x^d, $d\in\mathcal{D}$, durch die Gleichungen (3) voneinander abhängen. Der Digraph $D=(V,A)$ wird also von $|\mathcal{D}|$ verschiedenen Gütern durchströmt, wobei die Nebenbedingung besteht, daß für jede Fahrgastfahrt t genau einer der $|G(t)|$ parallelen Bögen (t^-, t^+) durchströmt werden muß.

Die Abbildung 1 zeigt ein Zwei-Depot-Problem mit fünf Fahrgastfahrten im Ort-Zeit-Diagramm und eine zulässige Lösung. Abbildung 2 zeigt die dazu gehörenden Digraphen (V, A) und (V_d, A_d).

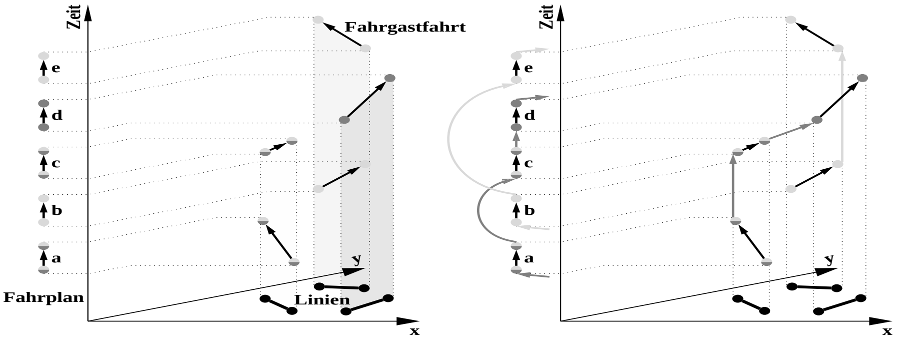

Abb. 1. Ort-Zeit-Diagramm und zulässige Lösung.

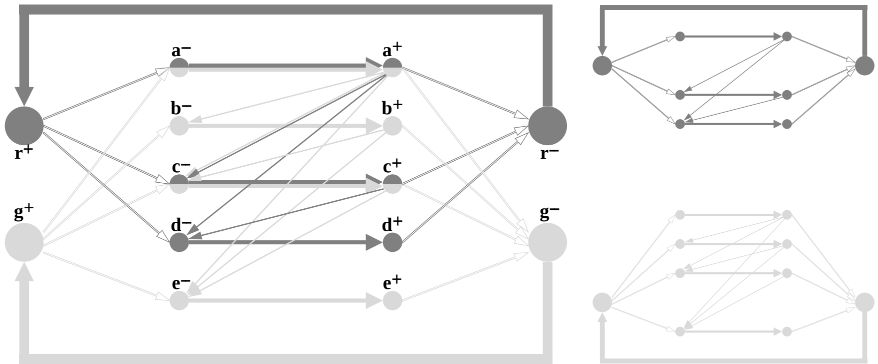

Abb. 2. Digraphen (V, A) und (V_d, A_d) zu $\mathcal{D} = \{$r,g$\}$, $\mathcal{T} = \{$a,b,c,d,e$\}$.

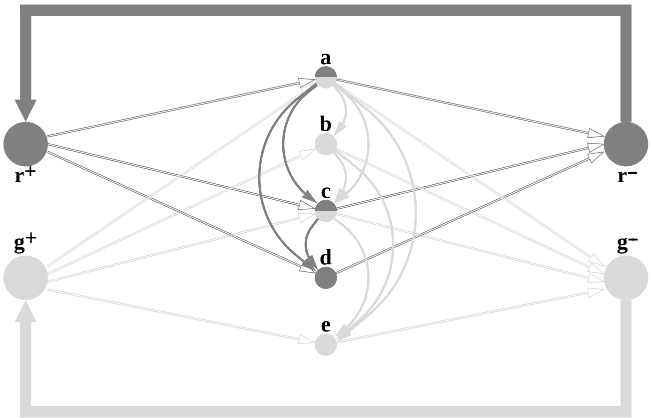

Abb. 3. "Geschrumpfter" Digraph (V', A').

Es ist offensichtlich, daß man die parallelen Bögen (t^-, t^+) zwischen dem Anfangs- und dem Endhaltepunkt einer Fahrgastfahrt t nicht benötigt und man daher die zugehörigen Variablen eliminieren kann (Grund: "Was in t^- hineinfließt, fließt wegen (2) aus t^+ wieder hinaus"). Graphentheoretisch ersetzt man für jede Fahrgastfahrt t die Knoten t^- und t^+ durch einen neuen Knoten t und entfernt alle Bögen zwischen t^- und t^+. Jeder Bogen mit Endknoten t^- erhält als neuen Endknoten den Knoten t, und analog beginnt jeder Bogen mit Anfangsnknoten t^+ nunmehr in t. Der so definierte neue Digraph $D' := (V', A')$ (mit $V' := \mathcal{D}^+ \cup \mathcal{D}^- \cup \mathcal{T}$, $A'_d := A_d^{\text{leer}} \cup \{(d^-, d^+)\}$ und $A' := \dot{\bigcup}_{d \in \mathcal{D}} A'_d$) entsteht also durch Schrumpfen der Knotenmengen $\{t^-, t^+\}$, $t \in \mathcal{T}$, und Entfernen der dabei auftretenden Schlingen. Abbildung 3 zeigt den geschrumpften Digraphen (V', A') zu dem Beispiel der Abb. 1 und Abb. 2.

Durch geeignete Kombination der Gleichungen (2) und (3) kann man beweisen, daß die Gleichungen (3) nach Elimination aller Variablen der Form $x^d_{(t^-,t^+)}$ äquivalent zu den folgenden Gleichungen sind:

$$\sum_{d \in \mathcal{D}} \sum_{a \in \delta^-(t) \cap A'_d} x^d_a = 1 \quad \forall\, t \in \mathcal{T}. \tag{5}$$

Jede zulässige (optimale) Lösung des linearen ganzzahligen Programms

$$\min \sum_{d \in \mathcal{D}} \sum_{a \in A'_d} c^d_a\, x^d_a \tag{6}$$

$$\sum_{d \in \mathcal{D}} \sum_{a \in \delta^-(t) \cap A'_d} x^d_a = 1 \quad \forall\, t \in \mathcal{T}, \tag{7}$$

$$\sum_{a \in \delta^-(v) \cap A'_d} x^d_a - \sum_{a \in \delta^+(v) \cap A'_d} x^d_a = 0 \quad \forall\, v \in V' \;\; \forall\, d \in \mathcal{D}, \tag{8}$$

$$l^d_a \leqslant x^d_a \leqslant u^d_a \quad \forall\, a \in A' \;\; \forall\, d \in \mathcal{D}, \tag{9}$$

$$x \text{ ganzzahlig} \tag{10}$$

definiert somit eine zulässige (optimale) Lösung des Umlaufplanungsproblems und umgekehrt. Man kann auch noch die zu den Rückführungsbögen gehörenden Variablen und die redundanten Kapazitätsschranken $x^d_a \leqslant 1$, für alle $a \in A_d^{\text{leer}}$ und $d \in \mathcal{D}$, eliminieren, aber darauf gehen wir nicht ein. In [FHW] und [BCP] werden ähnliche Modelle zu (6-10) vorgeschlagen; diese verwenden ebenfalls Depotgruppen.

4 Heuristiken

Viele Verkehrsbetriebe haben manuelle und/oder computerunterstützte Verfahren zur Bestimmung brauchbarer Umlaufpläne entwickelt. Wir kennen keinen Verkehrsbetrieb, der das von uns in den Abschnitten 2 und 3 dargestellte mathematische Modell (oder ein dazu äquivalentes) benutzt und somit in der Lage wäre, Umlaufpläne mit minimaler Fahrzeugzahl oder mit minimalen Kosten zu berechnen. Einige Software-Firmen und Software-Abteilungen von Verkehrsbetrieben vertreiben Programme zur Berechnung günstiger Umlaufpläne. Die dabei benutzten Methoden beruhen auf relativ ähnlichen heuristischen Prinzipien.

Betrachtet man das von uns vorgestellte Modell, so bietet sich eine heuristische Vorgehensweise direkt an. Es ist bekannt, daß das Mehrgüterflußproblem *NP*-schwer ist, aber Probleme mit einem Gut, d. h. mit einem Depot, in polynomialer Zeit lösbar sind. Was liegt näher, als alle Depots zu einem Depot zusammenzufassen und das so vereinfachte Problem zu lösen. Bei diesem Vorgehen entstehen einige Umläufe bzw. Umlaufketten, die nicht "depot-rein" gefahren werden. Man schließt daher eine heuristische Nachbearbeitung der Lösung an, bei der Umläufe in depot-reine Stücke zerlegt und diese anschließend depot-rein verheftet werden. Dadurch erhält man für jedes Depot eine Menge von Umläufen, die von diesem Depot bedient werden, und somit implizit eine Zuweisung von Fahrgastfahrten zu Depots. Man kann dann anschließend die einem Depot zugeordneten Fahrgastfahrten als ein Ein-Depot-Umlaufplanungsproblem auffassen und mit einem Minimal-Kosten-Fluß-Algorithmus optimal lösen. Diese Vorgehensweise kann man kurz mit dem Schlagwort *Schedule – Cluster – Schedule* beschreiben. Wir zeigen in Abschnitt 7, daß derartige Heuristiken Umlaufpläne liefern, die sowohl bez. der Fahrzeuganzahl als auch der Kosten recht nahe am Optimum liegen können.

Die derzeit in der Praxis benutzten Heuristiken unterscheiden sich in der Art, wie die einzelnen der hier skizzierten Schritte ausgeführt werden. Vielfach wird der letzte Schedule-Schritt weggelassen; man zeigt sich mit der heuristischen Verknüpfung zufrieden. Bei anderen Ansätzen werden die Cluster auf eine andere Weise bestimmt. Es werden z. B. rein heuristische Zuweisungen zu Depots vorgenommen, oder es werden Assignment- oder Transportalgorithmen benutzt um eine Clusterung zu erhalten (einen Überblick dazu geben z. B. [DP] oder [DDSS]). Die letztere Methodik wird bei HOT II verwendet, siehe [DMS], [DM] und [Mo].

Weitere Heuristiken des Schedule–Cluster–Schedule Typs sind in [BCP], [La], [BCG] und [MP] beschrieben. Die numerischen Testrechnungen dieser Publikationen beschränken sich jedoch auf Beispiele mit maximal 600 Fahrgastfahrten, meistens sind es noch deutlich weniger.

5 Ein exakter Lösungsansatz

In Abschnitt 3 haben wir gezeigt, daß das Umlaufplanungsproblem durch das ganzzahlige Programm (6-10) gelöst werden kann. Umlaufplanungsprobleme von großen Verkehrsbetrieben führen zu derartigen ganzzahligen Programmen mit mehreren Millionen Variablen. Aufgrund dieser enormen Größenordnungen sind nur wenige Versuche unternommen wurden, reale Problembeispiele exakt zu lösen. Forbes et al. [FHW] berichten von praktischen Problemen mit drei Depots und 6.500 Fahrgastfahrten. Mit ihrem Algorithmus und den ihnen zur Verfügung stehenden Rechnern konnten jedoch nur Teilprobleme mit höchstens 600 Fahrgastfahrten und rund 90.000 Leerfahrten gelöst werden. Ribeiro und Soumis [RS] lösen mit Column-Generation-Technik (Dantzig-Wolfe-Dekomposition) zufällig erzeugte Umlaufplanungsprobleme mit bis zu 10 Depots und bis zu 300 Fahrgastfahrten.

Unsere Idee war, LP-Techniken, die sich in letzter Zeit gerade bei der Lösung sehr großer kombinatorischer Optimierungsprobleme bewährt haben, in Verbindung mit polyedertheoretischen Überlegungen einzusetzen. Wir hatten die Hoffnung, daß reale Umlaufplanungsprobleme vielleicht doch nicht ganz so schwierig sind, wie es komplexitätstheoretische Überlegungen erwarten lassen. Diese Hoffnung hat sich bestätigt.

Die Idee des "polyedrischen Ansatzes" ist es, die konvexe Hülle MDVS(D') der zulässigen Punkte von (6-10) zu betrachten. MDVS(D') ist ein Polytop, dessen Ecken den Lösungen des durch den Digraphen D' definierten Umlaufplanungsproblem entsprechen. Aus der Polyedertheorie ist bekannt, daß es ein System von (i. a. sehr vielen) Ungleichungen $Bx \leqslant b$ und Gleichungen $Ex = e$ gibt, so daß gilt: $\mathrm{MDVS}(D') = \{x \in \mathbb{R}^A \mid Bx \leqslant b,\ Ex = e\}$.

Diese (theoretische) Konstruktion erlaubt es, das Umlaufplanungsproblem als lineares Programm der Form $\min\{c^{\mathrm{T}}x \mid Ex = e, Bx \leqslant b\}$ zu lösen. In der Praxis kennt man nur einen winzigen Bruchteil aller notwendigen Ungleichungen. Aber selbst dieser Bruchteil besteht schon aus einer Zahl von Ungleichungen, die exponentiell in der Knotenzahl von D' ist. Man geht daher wie folgt vor: Zunächst bestimmt man eine "vernünftige" Teilmenge der Ungleichungen, sagen wir das Teilsystem $B'x \leqslant b'$ von $Bx \leqslant b$, und löst das lineare Programm $\min\{c^{\mathrm{T}}x \mid Ex = e, B'x \leqslant b'\}$. Entspricht die Optimallösung einem Umlaufplan, so ist man fertig. Andernfalls wird versucht, aus der Menge der bisher nicht berücksichtigten Ungleichungen einige zu finden, die die gegenwärtige gebrochene Optimallösung abschneiden. Diese Ungleichungen werden *Schnittebenen* genannt. Die Schnittebenen werden dann zum gegenwärtigen linearen Programm hinzugefügt, und es wird iteriert, bis entweder ein optimaler Umlaufplan gefunden ist oder die gegenwärtige optimale Lösung eines linearen Teil-Programms nicht mehr abgeschnitten werden kann. In diesem Fall wird ein Branch-and-Bound Verfahren angeschlossen.

Wir haben das Polytop MDVS(D') untersucht und verschiedene Klassen von Facetten gefunden. Die Darstellung dieser Ergebnisse ist technisch aufwendig, weswegen wir hier nicht darauf eingehen.

Unsere Rechenexperimente haben gezeigt, daß die Lösung der LP-Relaxierung von (6-10), die durch Weglassen der Ganzahligkeitsbedingung (10) entsteht, für reale Probleme einen Optimalwert ergibt, der nahe am Optimalwert des Umlaufplanungsproblems liegt. In der Regel lieferte die LP-Relaxierung die minimale Fahrzeuganzahl, und die optimalen Kosten liegen nie mehr als 1% über den Kosten der optimalen LP-Lösung. Die Schnittebenen wurden benötigt, um diese kleine Lücke zu schließen und aus den gebrochenen optimalen LP-Lösungen ganzzahlige Optimallösungen zu erzeugen.

Wirklich schwierig war es jedoch, die LP-Relaxierungen selbst zu lösen. Immerhin handelt es sich hier um lineare Programme, die bei realen Datensätzen bis zu 50 Millionen Variablen und 60.000 Gleichungen umfaßen. Hierfür haben wir spezielle Techniken entwickelt, die wir in Abschnitt 6 beschreiben.

6 Implementierungsdetails

Bei LP-Problemen mit mehreren Millionen Variablen ist allein aus Speicherplatzgründen von vornherein klar, daß die LP's nicht vollständig erzeugt und einem Löser übergeben werden können. In unserem Fall bot es sich an, alle Gleichungen und Ungleichungen von (6-10) zu berücksichtigen und Variablen sukzessive durch "column generation" einzubeziehen. Nach verschiedenen Experimenten hat sich der folgende Weg als erfolgreich erwiesen.

Im ersten Schritt wird eine heuristische Lösung (siehe Abschnitt 4) des vorliegenden Problems bestimmt. Danach werden für das erste lineare Programm all diejenigen Spalten von Variablen generiert, deren Leerfahrten in der heuristisch gefundenen Lösung benutzt werden. Das so erzeugte LP wird dann mit CPLEXMACH, einer Beta-Version von CPLEX4.0 [Cp] gelöst. Anschließend beginnt die Phase A des Lösungsprozeßes.

Phase A: Wir überprüfen die reduzierten Kosten aller Nichtbasis-Variablen und löschen diejenigen Variablen aus dem gegenwärtigen LP, deren reduzierte Kosten eine gewisse (parametergesteuerte) Schranke überschreiten (heuristische Reduktion der LP-Größe). Nacheinander werden die folgenden drei Spaltengenerierungsalgorithmen aufgerufen:

SG1: Die Gleichungen (7) werden aus dem LP gestrichen und über einen Lagrange-Relaxierungsansatz zur Zielfunktion hinzugefügt. Die Lagrange-Multiplikatoren erhalten die Werte der Dualvariablen zu (7) des zuletzt gelösten LP's. Das verbleibende LP dekomponiert in $|\mathcal{D}|$ unabhängige Minimal-Kosten-Fluß-Probleme. Wir lösen diese mit dem Netzwerk-Simplex-Code MCF ([Lo2]). Alle Variablen, die in der optimalen Lösung eines der Minimal-Kosten-Fluß-Probleme einen positiven Flußwert aufweisen, werden zum gegenwärtigen LP hinzugefügt, falls sie dort nicht schon vorhanden sind.

SG2: Durch geeignete Kombination der Gleichungen (7) und (8) erzeugen wir ein (redundantes) Gleichungssystem der Form

$$-\sum_{d \in \mathcal{D}} \sum_{a \in \delta^+(t) \cap A'_d} x_a^d \quad = \quad -1 \quad \forall\, t \in \mathcal{T}. \tag{11}$$

Die Gleichungen (8) und die Schranken für die Depotkapazität von (9) werden aus dem LP gestrichen und über einen Lagrange-Ansatz zur Zielfunktion hinzugefügt. Die Lagrange-Multiplikatoren erhalten auch hier die Werte der zugehörigen Dualvariablen zu den Gleichungen (8) und Ungleichungen (9) des zuletzt gelösten LP's. Die verbleibenden Gleichungen (6), (7) und (11), die Ungleichungen $x \geqslant 0$ und die neue (Lagrange-) Zielfunktion bilden ein (riesiges) Minimal-Kosten-Fluß-Problem, welches wiederum mit dem Netzwerk-Simplex-Code MCF gelöst wird. Analog zu SG2 werden dann alle Variablen, die in der optimalen Lösung des Minimal-Kosten-Fluß-Problems einen positiven Flußwert besitzen und sich nicht bereits im aktuellen LP befinden, zum LP hinzugefügt.

SG3: Die reduzierten Kosten aller nicht im gegenwärtigen LP vorhanden Variablen werden bestimmt. Die Variablen mit den geringsten reduzierten Kosten werden (parametergesteuert) zum LP hinzugefügt.

Nach der Spaltenerzeugung wird das veränderte LP mit dem dualen Simplexverfahren und einem vorgeschalteten Preprocessing gelöst. Ist der Minimalwert des neuen LP's "deutlich niedriger" als der des vorhergehenden LP's, so wird die Phase A wiederholt, andernfalls wird zur Phase B gesprungen.

Phase B: Wie in Phase A werden zuerst Variablen aus dem gegenwärtigen LP entfernt. Anschließend wird lediglich SG3 aufgerufen. Das so veränderte LP wird in dieser Phase mit dem primalen Simplexalgorithmus gelöst, wobei die optimale Basis des letzten LP's als Startbasis verwendet wird. Die Phase B wird solange wiederholt, bis in SG3 die globale Optimalität der Optimallösung des gegenwärtigen LP's nachgewiesen ist.

Die Spaltengenerierungstechnik SG3 is notwendig, um am Ende die globale Optimalität einer LP-Lösung mit reduziertem Variablensatz nachzuweisen. Allerdings haben praktische Experimente gezeigt, daß diese Methode allein LP's des vorliegenden Typs nicht löst. Es wird praktisch kein Fortschritt im Zielfunktionswert erreicht ("stalling"), außerdem lassen sich kaum Variablen aufgrund ihrer reduzierten Kosten aus dem LP entfernen.

Die beiden Spaltengenerierungstechniken SG1 und SG2 bringen eine globalere Sichtweise als SG3: Es wird nicht nur eine einzelne Variable für sich allein aufgrund ihrer reduzierten Kosten beurteilt, sie wird vielmehr im Zusammenspiel mit den übrigen Variablen bewertet. Auf diese Weise werden ganze Umläufe, die natürlich in SG2 nicht notwendig depot-rein sind, generiert. Hierbei können durchaus auch Variablen zum LP hinzugefügt werden, deren reduzierte Kosten positiv sind, also lokal gesehen keine Verbesserung des Zielfunktionswerts versprechen, jedoch zur Komplettierung ansonsten günstiger Umläufe benötigt werden.

Es ist aus Platzgründen unmöglich, alle Details der gesamten Lösungsmethode zu beschreiben. Wir wollen jedoch noch erwähnen, daß wir heuristisch versuchen, aus jeder LP-Lösung eine zulässige Lösung zu generieren. Dies führt bei unseren Problemen zu deutlichen Verbesserungen der oberen Schranke, was das anschließende Branch&Cut-Verfahren erheblich beschleunigt.

7 Rechenergebnisse

Die oben beschriebenen Methoden wurden insbesondere auf den in Tab. 1 skizzierten realen Datensätzen der Hamburger Hochbahn AG, der Berliner Verkehrsbetriebe und der Verkehrsbetriebe Hamburg-Holstein AG getestet:

Tabelle 1. Mehr-Depot-Probleme.

Verkehrsbetrieb	Depots	Fahrgastfahrten	Leerfahrten
Hamburg	40	16.239	15.100.000
Hamburg 1	12	8.563	11.400.000
Hamburg 2	9	1.834	1.100.000
Hamburg 3	2	791	200.000
Hamburg 4	2	238	24.000
Hamburg 5	2	1.461	620.000
Hamburg 6	2	2.283	1.700.000
Hamburg 7	2	341	36.000
Berlin	49	21.003	50.000.000
Hamburg-Holstein	4	3.413	3.700.000

Hamburger Hochbahn AG: In Hamburg werden zur Zeit 14 Betriebshöfe (z. T. Fremdunternehmen) mit insgesamt 9 verschiedenen Fahrzeugtypen geplant. Werden die einzelnen Fahrzeugtypen, die jeweils in einem Betriebshof stationiert sind, als ein Depot definiert, so sind in Hamburg 40 Depots zu unterscheiden. Der Datensatz umfaßt 16.239 Fahrgastfahrten, die für alle Depots zusammen durch rund 15,1 Millionen Leerfahrten verknüpft werden können. Das Gesamtproblem dekomponiert in ein 12-Depot-Problem, ein 9-Depot-Problem, fünf 2-Depot-Probleme und neun kleinen Ein-Depot-Probleme mit insgesamt 728 Fahrgastfahrten.

Berliner Verkehrsbetriebe: Die BVG unterhält zur Zeit 9 Betriebshöfe mit 10 verschiedenen Fahrzeugtypen. Insgesamt ergeben sich für die BVG 49 Depots. In Berlin sind an einem normalen Wochentag rund 28.000 Fahrgastfahrten zu erledigen. Durch Vergabe von Fahrgastfahrten bzw. ganzer Linien an Fremdunternehmen reduziert sich das Problem auf 21.003 Fahrgastfahrten und rund 50 Millionen Leerfahrten.

Verkehrsbetriebe Hamburg-Holstein AG: Hier werden 18 Betriebshöfe (z. T. Fremdunternehmen) geplant. Eine Unterteilung in Fahrzeugtypen wurde in unseren Testdaten nicht vorgenommen, jeder Betriebshof definiert deshalb ein Depot. Die 5.603 Fahrgastfahrten können mit rund 4,2 Millionen

Leerfahrten verknüpft werden. Auch dieses Problem dekomponiert in ein 4-Depot-Problem mit 3.413 Fahrgast- und 3,7 Millionen Leerfahrten und in 14 leinere Ein-Depot-Probleme mit bis zu 613 Fahrgast- und ca. 150.000 Leerfahrten.

Wir gehen auf die Ein-Depot-Probleme nicht weiter ein, da diese wie bereits erwähnt polynomial lösbar sind. Tabelle 2 enthält die Zielfunktionswerte der von uns mit dem Branch&Cut-Verfahren berechneten Optimallösungen sowie die besten Werte der heuristisch bestimmten Lösungen. Die benötigten Rechenzeiten finden sich in Tab. 3.

Tabelle 2. Fahrzeugbedarfe mit Gewichtungen.

Verkehrsbetrieb	Optimalwerte		Heuristische Lösung	
	Fahrzeuge	Gewichtung	Fahrzeuge	Gewichtung
Hamburg (alle Depots)	812	117540	833	113272
Hamburg 1	432	63074	449	60918
Hamburg 2	103	14702	104	15032
Hamburg 3	39	5085	40	4937
Hamburg 4	6	1205	6	1205
Hamburg 5	62	10920	63	11167
Hamburg 6	111	14330	111	14444
Hamburg 7	15	2655	16	2530
Berlin	1157	194202	1166	197849
Hamburg-Holstein	201	26405	202	28052

Tabelle 3. Rechenzeiten in Stunden:Minuten auf SUN SPARCStation 20-71.

Verkehrsbetrieb	optimal	fahrzeug-minimal	heuristisch
Hamburg 1	56:47	50:10	0:10
Hamburg 2	4:07	2:52	0:05
Hamburg 3	0:04	0:03	0:01
Hamburg 4	0:01	0:01	0:01
Hamburg 5	0:08	0:05	0:03
Hamburg 6	0:06	0:01	0:01
Hamburg 7	0:01	0:01	0:01
Berlin	41:00	5:00	0:27
Hamburg-Holstein	1:55	0:33	0:03

Das von uns entwickelte Branch&Cut-Verfahren ist in der Lage, in akzeptablen Rechenzeiten Optimallösungen dieser wirklich riesigen Umlaufplanungsprobleme zu bestimmen. Die Heuristiken liefern in wenigen Minuten sehr gute Lösungen, deren Güte durch das Schnittebenenverfahren in kurzer Zeit abgeschätzt werden kann. Diese und andere Vergleichsrechnungen haben gezeigt, daß das erzielbare Einsparungspotential von den im Einsatz befindlichen Planungsmethoden abängt. Bei computergestützter manueller Planung können z. T. rund 10 % der Fahrzeuge eingespart werden. Die besten Heuristiken der Praxis (hierzu gehört HOT II) liefern sehr gute Lösungen. Die von uns berechneten Einsparungspotentiale lagen für die Fahrzeugzahl im Bereich von 3 % und bei den Kosten im Bereich von bis zu 10 %.

Literatur

[BCG] A. A. Bertossi, P. Carraresi und G. Gallo. On some matching problems arising in vehicle scheduling models. *Networks*, 17:271–181, 1987.

[BCP] I. Branco, A. Costa und J. M. P. Paixão. Vehicle scheduling problem with multiple type of vehicles and a single depot. In Daduna et al. [DBP].

[BMMN] M. O. Ball, T. L. Magnanti, C. L. Monma und G. L. Nemhauser, editors. *Network Routing*, volume 8 of *Handbooks in Operations Research and Management Science*. Elsevier Science B.V., Amsterdam, 1995.

[Cp] CPLEX Optimization, Inc., Suite 279, 930 Tahoe Blvd., Bldg 802, Incline Village, NV 89451, USA. *Using the CPLEX Callable Library*, 1995. URL: `http://www.cplex.com/`.

[DBP] J. R. Daduna, I. Branco und J. M. P. Paixão, editors. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, 1995.

[DDSS] J. Desrosiers, Y. Dumas, M. M. Solomon und F. Soumis. *Time Constrained Routing and Scheduling*. In Ball et al. [BMMN], Kapitel 2.

[DM] J. R. Daduna und M. Mojsilovic. Computer-aided vehicle and duty scheduling using the HOT programme system. In Daduna und Wren [DW].

[DMS] J. R. Daduna, M. Mojsilovic und P. Schütze. Practical experiences using an interactive optimization procedure for vehicle scheduling. In Du und Pardalos [DP], Seiten 37–52.

[DP] J. R. Daduna und J. M. P. Paixão. Vehicle scheduling for public mass transit – an overview. In Daduna et al. [DBP].

[DP] D.-Z. Du und P. M. Pardalos, editors. *Network Optimization Problems: Algorithms, Applications and Complexity*, volume 2 of *Series on Applied Mathematics*, Singapore, New York, London, 1993. World Scientific Publishing Co. Pte. Ltd.

[DR] M. Desrochers und J.-M. Rousseau, editors. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, 1992.

[DW] J. R. Daduna und A. Wren, editors. *Computer-Aided Transit Scheduling*, Lecture Notes in Economics and Mathematical Systems. Springer Verlag, 1988.

[FHW] M. A. Forbes, J. N. Holt und A. M. Watts. An exact algorithm for multiple depot bus scheduling. *European Journal of Operational Research*, 72(1):115–124, 1994.

[FP] R. Freling und J. M. P. Paixão. Vehicle scheduling with time constraint. In Daduna et al. [DBP].

[La] A. Lamatsch. *Wagenumlaufplanung bei begrenzten Betriebshofkapazitäten*. Dissertation, Universität Fridericiana zu Karlsruhe (TH), 1988.

[Lo1] A. Löbel. *Optimal Vehicle Scheduling in Public Transit*. Dissertation, Technische Universität Berlin, erscheint 1996.

[Lo2] A. Löbel. Solving large-scale real-world minimum-cost flow problems by a network simplex method. Preprint SC 96-7, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 1996. URL:
`ftp://ftp.zib.de/pub/zib-publications/reports/SC-96-07.ps.Z`.

[LS] A. Löbel und U. Strubbe. Wagenumlaufoptimierung – Methodischer Ansatz und praktische Anwendung. In *Heureka '96: Optimierung in Verkehr und Transport*, Seiten 341–355. Forschungsgesellschaft für Straßen- und Verkehrswesen, Köln, 1996. URL:
`ftp://ftp.zib.de/pub/zib-publications/reports/SC-95-38.ps.Z`.

[Mo] M. Mojsilovic. Verfahren für die Bildung von Fahrzeugumläufen, Dienstplänen und Dienstreihenfolgeplänen. In *Heureka 83 – Optimierung in Transport und Verkehr*, Seiten 178–191. Forschungsgesellschaft für Straßen- und Verkehrswesen, Köln, 1983.

[MP] M. Mesquita und J. M. P. Paixão. Multiple depot vehicle scheduling problem: A new heuristic based on quasi-assignment algorithms. In Desrochers und Rousseau [DR].

[PS] P. Petzold und P. Schütze. Integrated data processing for public transport in Hamburg. In Daduna et al. [DBP].

[RS] C. C. Ribeiro und F. Soumis. A column generation approach to the multiple-depot vehicle scheduling problem. *Operations Research*, 42(2):41–52, 1994.

[SV] P. Schütze und M. Völker. Recent developments of HOT II. In Daduna et al. [DBP].