From Vehicle Scheduling to Vehicle Routing:
Problem structures, (dis-)similarities, complexities, practical issues, and efficient solution by “Merged Network Flows” Math Models

Part B: Modelling towards Large Scale Optimization
→ Part A: Problem Classes and Research Road Map (submitted as a separate contribution)

Taïeb Mellouli, Prof. Dr.
Department of Business Information Systems and Operations Research,
Martin-Luther-University Halle-Wittenberg, Universitätsring 3, 06108 Halle Saale, Germany
mellouli@wiwi.uni-halle.de

keywords:

Extended Abstract (for Part B: Modelling towards large scale optimization)

1. Overview and focus on efficiently solvable mathematical models

The main focus of this paper is to show that mathematical modeling and optimization can deliver good results in practice in spite of the high complexity of the underlying vehicle scheduling and routing problems. This is achieved by a modeling exploiting inherent characteristics of problem structures, by using high-performance mathematical optimization solvers, and by integrating special techniques.

The Vehicle Scheduling and Vehicle Routing problems (VSP and VRP) are the subject of Part A of our research work where problem structures and their interrelations are discussed. Based on author’s own research in both bus scheduling and rotation planning in railways, it can be noticed, that state-of-the-art methods for large scale instances of vehicle scheduling with its practical variants are based on mathematical programming (exact optimization). On the other side and based on the vast literature on vehicle routing and their relevant variants, heuristic and metaheuristic approaches are mostly used for practical large-scale instances in logistics for transportation of goods. So, thinking of VSP as a complex variant of VRP can be misleading and leads to developing a heuristic solution for a problem which can be optimally solved by mathematical programming for large-scale practical instances, as will be shown in section 3 for the multiple vehicle type and multiple depot vehicle scheduling problem. Our research map suggests that it is beneficial to transmit modeling knowhow of mathematical optimization from vehicle scheduling to vehicle routing and not the other way around, as in the major literature.

This Part of our research work shows advanced modeling and optimization techniques developed by the author allowing for an almost direct solution by MIP-Solvers for middle and large-scale instances of VSP with its practically relevant variants in both bus transport and railways. Furthermore, we show that understanding the basic problem structure is crucial in order to design successful, efficiently solvable network-flow based mathematical models for VSP. In a second step in this paper, we show how this line of thinking can further help to design efficiently solvable mathematical models also for the harder case of VRP variants. As explained in Part 1, subsection 3.B, we concentrate on the vehicle routing problem with restricted mixing of deliveries and pickups, as it is a practically relevant problem in logistics saving operational costs in terms of route lengths by combining pickups before returning empty from delivery activities, and second because our analysis of basic structure on this problem has led to a better understanding of the relation between VPR, VPRB and VPRDP which can all be solved by the same proposed mathematical model by a mere change of one new parameter τ being the number of DP-turns within a mixed delivery and pickup tour (3.B in Part A, and in section 4 below).

We emphasized in the last paragraph, that our methodology is based on designing mathematical models which can be directly solved by MIP-Solvers for middle-sized instances without resort to complex
technique such as column generation, Lagrange relaxation or the like. We do use other techniques for solving large-scale instances, e.g., **iteratively fixing flows from the efficiently computed solution of LP relaxation**, however our models have the properties that they are more performant when used without further complicated techniques. Our approach is based on **exploiting the problem structure** (see next section) already in the design of the underlying network of the developed network-flow based mathematical models. Further, our models are enhanced by techniques for **aggregating, coupling, or generally, merging flows**, which can be seen in some cases as **network reduction techniques**, because flows are enforced to take some **unified paths** within the network, instead of individual or separated paths. This has tremendous effect on reducing symmetries, thus avoiding numerical problems when solved by MIP-Solvers. The underlying modeling techniques shown in section 3 for VSP extensions and in section 4 for VPRDP can be subsumed under the unifying term “**Merged Network Flows**”.

2. **Exploiting the problem structures in modeling**

Our mathematical modeling approach is not classical, but **exploits problem structures in a higher extent**. Therefore, we do not start with a standard mathematical formulation and further complicate it by logical constraints and the like which can be written as perhaps nice formulas using big-M technique etc., because one gets only a correct model, but the model is in most cases not efficiently solvable for practically relevant instances (I personally see this in many research contributions). Instead of this, we invest more in thoroughly analyzing the complex kernel of the problem and try to **unveil the internal structure of the problem**, then exploit this problem structure in designing a network or a network with flows in order to model the complex dependencies within the network. The more practical constraints are encoded into the flow network the less extra constraints are needed. Our experience suggests that with a **high encoding rate of problem structure** dependencies and practical constraints within the model (flow) networks, the mathematical model based on flows within these networks is likely to be efficiently solved by MIP solvers for larger instances. Based on the research road map of Part 1, let us summarize our findings about internal structures and inter-dependencies of the problems VSP and rotation building for public transport, on the one side, and the VRPDP, especially the vehicle routing problem with restricted mixing of deliveries and pickups, on the other side, in the following table:

<table>
<thead>
<tr>
<th>Problems</th>
<th>VRP</th>
<th>VRPB</th>
<th>VRPDP</th>
<th>PDP</th>
<th>DARP</th>
<th>VSP(-DH)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Icons”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>XXX</td>
</tr>
<tr>
<td>Schema</td>
<td>o-D</td>
<td>o-M-o</td>
<td>o-M-o</td>
<td>1-1</td>
<td>m</td>
<td>m* (P-D)</td>
</tr>
<tr>
<td>Transport</td>
<td>depot goods</td>
<td>depot goods</td>
<td>depot goods</td>
<td>PD-goods</td>
<td>People</td>
<td>1000’s (!!)</td>
</tr>
<tr>
<td>Input</td>
<td>Customers</td>
<td>D/P-nodes</td>
<td>D/P-nodes</td>
<td>P-D-requests</td>
<td>dialed Rides</td>
<td>timetabled Trips</td>
</tr>
<tr>
<td>Instances in Practice</td>
<td>100’s</td>
<td>100’s</td>
<td>100’s</td>
<td>~50-100</td>
<td>~30-60</td>
<td>1000’s (!!)</td>
</tr>
<tr>
<td>TW-basic</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>~/medium</td>
<td>tight</td>
<td>fixed times</td>
</tr>
<tr>
<td>TW-extension</td>
<td>large</td>
<td>large</td>
<td>large</td>
<td>-</td>
<td>-</td>
<td>tight</td>
</tr>
<tr>
<td>Precedence</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>P→D</td>
<td>P→D</td>
<td>S→E</td>
</tr>
<tr>
<td>Intra-(PD)-request</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>possible</td>
<td>only within bounds</td>
<td>SSEE not possible</td>
</tr>
<tr>
<td>INTER-Compatibility</td>
<td>oD,,Do or oP,,Po (distances)</td>
<td>D,,D’&lt;P,,P’ no mix (distances)</td>
<td>D,,D’&lt;P,,P’ preferable (distances)</td>
<td>P’D’&lt;P”D” preferable (distances)</td>
<td>P’D’&lt;P”D” strongly pref. (distances)</td>
<td>S’E’S”E” + Intertrip-DH (.t) pot. DeadHeads</td>
</tr>
<tr>
<td>Constraints (Wasan/Nagy)</td>
<td>-</td>
<td>no mixing of D-&amp;P-loads</td>
<td>restricted mixing</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Tour Quality (Melloul)</td>
<td>τ = 0</td>
<td>τ = 1</td>
<td>restrict τ</td>
<td>#D-P-turns</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Investigating internal structure and inter-dependencies of vehicle routing and scheduling problems
One feature of our study of internal structure of vehicle scheduling and vehicle routing problem is unveiling the tight relationship of VSP and DARP. In light of the fact that efficient mathematical models exist for VSP with deadheading for instances with thousands of trips (see next section), the VSP can be extended with tight TWs and adapted to solve some DARP basic cases where practically relevant instances have only 30 to 60 rides (pickup-delivery) requests. For each tight TW, one can generate some tens of alternative deadhead trips to each compatible next request instead of updating or linking times at nodes by big-M-constraints over arcs (making the model not efficiently solvable). So there is a real opportunity for transferring optimization knowhow from VSP to DARP.

A second feature is the clear distinction between precedence intra-PD-request, i.e. between pickup and corresponding delivery of same PD-request in PDP, and inter-compatibility relation between different PD-requests or rides. The precedence Intra-Request is meant between P und D of same request and this precedence does not play any role in VSPDP and this is in line with our insistence of the different characters of the PDPs, from one side, and VRPs (including those having all deliveries from depot and all pickups to depot). That is, VRPDP is different in character from PDPs and our table places PDP and DARP more in the side of VSP than VRP. Our analysis of VRPDP makes it more linked to VPR through our discussion in Part A (loading operations relevant for practical vehicle routing are really apparent and influences routing only in VPRDP, not in basic VPR) and through the analysis of tour quality for routings with mixed deliveries and pickups (see section 4).

In the sequel, efficiently solvable mathematical models are provided for the problems highlighted in the table. The methodology “Merged Network Flows” possessing the aforementioned advantages is better explicated by concrete vehicle scheduling (section 3) and routing problems (section 4).

3. Merged network flows for VSP: LFM-technique for merging potential inter-trip deadheads

The basic vehicle scheduling problem builds cost optimal rotations for vehicles (busses, trains) for a given timetable of scheduled trips. As indicated in the table above, the instances in practice may have thousands of trips, thus the trip-as-node networks (figure 2, part A) possessing a quadratic number of arcs corresponding to all potential links between compatible trips don’t favor that the underlying flow-based mathematical models perform well for practical VSP extensions. These inter-trip compatibility arcs can be direct connections within the same station (“Anschluss” in German), if the end station of trip is equal to the start station of the next trip, or they are indirect, if the connection uses an unloaded trip to the next scheduled trip called deadhead (“Leerfahrt” in German for empty or unloaded trip).

In order to obtain an efficiently solvable mathematical model by MIP Solvers, connection lines CL (similar to time-lines in airline fleet assignment) are introduced into the rotation planning field in order to merge the flows of direct connections over waiting arcs within each station (see figure 1). This technique could be used for rotation planning where deadhead trips are of restricted use. We used this technique for building train rotations with maintenance routing by further merging flows of trains with same maintenance state and enable state transitions between CLs with different states (Mellouli 2001, at the conference CASPT 2000, Berlin, we acquired a project with a software provider for bus transit).

However, for bus scheduling and other railway problems with extensive use of deadhead trips, the basic model remains with quadratic number of indirect connections (see figure 2), so we developed the following technique in order to merge the flows of indirect connections (deadheads) between different stations, as well: Organizing these indirect connections in groups of matchings (M) from each station k of ending trips of CL(k) to another (any but fixed) station k’ of starting trips of CL(k’), we observed that only first matches (FM) suffice since a connecting deadhead trip to k’ can use the waiting arcs in k’ in order to reach each other connection trip starting later at k’ (see figure 3 in the middle: for each ending trip in k, there is at most one first match to k’ for each station k’ ≠ k). The quadratic number of
deadhead trips or matches M being \( O(N^2) \) could be thus reduced to only \( O(N^*K) \) of FM where N is the number of trips and K the number of stations \( (K << N) \). For a railway case study (Mellouli 2001-2002, presentation at Deutsche Bahn, Frankfurt), the number of matches (potential deadhead trips) of over 5 million reduces to 112 thousand (the number of trips \( N = 7.666 \) and the number of stations \( K=30 \)). The second stage of network reduction could be reached by observing that several trips ending at station \( k \) may have the same first match in station \( k' \), so the latest starting first match from \( k \) to \( k' \) is sufficient, since each former first match can be subsumed by going through waiting arcs in station \( k \), then taking the remaining latest first match (LFM) to \( k' \). The number of arcs for matches reduces from \( M = 5.168 \) million to only LFM = 25.5 thousands latest first matches (only 0.5% of Ms are LFM!!)

With the help of this LFM-technique (fig. 3), both extended versions of rotation planning for railways (multiple-type vehicles and wagon groups) and vehicle scheduling for bus transit (multiple depots and multiple vehicle type) could be solved for large-scale instances (practice projects at the Decision Support & Operations Research Lab, University of Paderborn 2000-2003). The expansion of the model regarding different vehicle types and different depots is a usual extension to a multi-commodity-like network flow model, where additional cover constraints ensure that each trip is served by a vehicle in one of the layers. Figure 4 shows this kind of network expansion, on the left hand side, and, on the right hand side, exemplified computational results for a 2.050 trip schedule of the city of Halle with 12 layers corresponding to pairs of depot and bus type (4 depot and 3 bus types) showing that the resulting models with more than 100 thousand variables can be solved within minutes of MIP solver’s time. Note, that larger and more complex models could be solved for rotation problems for railways with locomotive/carriage groups with shared capacities (Mellouli 2003, Mellouli & Suhl ATMOS 2004).

4. Merged network flows for vehicle routing problem with deliveries and pickups

In the last few years (2015-2017), we developed Merged Network Flows math models for VRPDP and for some basic PDPs. In the following, we show the techniques developed for the vehicle routing problem with restricted mixing of deliveries and pickups. For vehicle routing, flow models suffer of subtours and lot of symmetry (using the classical model as in figure 1 in Part A, where the X vehicle flow is governed by the combined flow balance and cover constraints \( \text{in}_X(i) = \text{out}_X(i) = 1 \) to each non-depot node i). One way to break subtours is to introduce a second type of commodity or load flow besides the vehicle flow and to couple them in an appropriate way. This load flow on one arc models the amount of load onboard of the vehicle traversing the route part corresponding to that arc. In order to get a feasible solution, the load commodity flows are coupled with the vehicle routes’ flow at each arc, logically, if there is a load greater than zero, a vehicle must be present \( (X[i,j] = 1) \). Two different load commodity flows are introduced for linehaul/delivery, which we denote by \( F[i,j] \), and for backhaul/pickup, denoted by \( Fb[i,j] \). So the coupling constraints are of the known big-M form \( F[i,j] + Fb[i,j] \leq C* X[i,j] \) for all arc \( (i,j) \), where big-M is set to C, the capacity of the vehicle.

The model of Wasan and Nagy (2014) consists of these coupled flows constraints which are further tightened for restricted mixing by \( F[i,j] + Fb[i,j] \leq (1-\gamma) * C * X[i,j] \) where \( \gamma \) represents the minimum percentage of free space (e.g., \( \gamma = 0.25 \) meaning 25%) required when a mixture of goods in onboard (to avoid the shuffling problem discussed in Part A). Nagy, Wasan, and Salhi (2013) remarked that the tightened constraints are only needed for DP-links (only B-to-L arcs). Although a better performing two-index instead a three-index model (one arc variable for homogeneous vehicles instead of several ones to each vehicle) is used in Wasan and Nagy (2014), only instances with up to 33 requests could be solved. In (Mellouli, OR 2015, Wien), we showed that by the following major revisions of this model, we could solve larger instances up to 100 requests, thereafter we integrated special techniques to solve larger instances up to 200 requests (OR 2017, Berlin), thus reaching better quality solutions than even the heuristic approach in Nagy, Wasan, and Salhi (2013).
The function of loads along a route for VRPB (with the strict rule “backhaul only after all linehuals”) is analyzed and it has been observed that only one D-P-arc is needed/used (see Figure 5). We introduce the notion of DP-turns generally for VRPPD and restrict their average number per tour by a parameter \( \tau \) (see Figure 6), resulting in a further quality indicator for delivery-and-pickup routes (in addition to minimum free space by restricted mixing restriction): \( \tau = 1 \) means that linehaul is always before backhaul (corresponding to VRPB, VRP with backhaul), \( \tau = 0 \) correspond to a version almost equal to basic VPR since the solution contains tours with only deliveries and tours with only pickups (exactly as solving basic VRP two times: for only the deliveries resulting in usual delivery-tours and for only the pickups resulting in pickup-tours in a collecting setting of VRP). The larger is \( \tau \) the more general is the VRP version towards VRPDP. We remarked, that a nearly optimal solution is reached in the majority of instances for a \( \tau \) value around 5 or less, that is, we could reduce the number of DP-turns without deteriorating the overall costs, but reaching a better quality of delivery and pickup tours with less problems of shuffling in load delivery operations caused by picked up loads onboard.

Further, we were convinced that recognizing route parts with linehaul loads and those with backhaul loads allows for an efficient modeling of restricted mixing (RM) by “streamlining” the LP solution towards real MIP solutions. To achieve this goal, we introduced load-type-flow indicator 0/1-variables on all arcs \( XF[i,j] \) and \( XFb[i,j] \) to logically indicate whether a vehicle contains linehaul and backhaul loads, respectively. These load-type-indicator flow can be coupled twice to the corresponding load flows by \( F[i,j] \leq C * XF[i,j] \) and \( Fb[i,j] \leq C * XFb[i,j] \) and to the main vehicle flow by \( XF[i,j] \leq X[i,j] \) and \( XFb[i,j] \leq X[i,j] \). Furthermore, we bind the vehicle-indicator flows along paths involving several arcs from depot to last delivery for \( XF \) and from first pickup to depot for \( XFb \). Whereas the normal X vehicle flow from depot to depot in governed by the flow balance constraints \( \text{in}_\text{flow}_X(i) = \text{out}_\text{flow}_X(i) (=1) \) for each non-depot node i, the indicator flows for linehaul and backhaul are open from one direction and governed by \( \text{in}_\text{flow}_XF(i) \geq \text{out}_\text{flow}_XF(i) \) since \( XF \) starts at depot and ends with last delivery and \( \text{in}_\text{flow}_XFb(i) \leq \text{out}_\text{flow}_XFb(i) \) since \( XFb \) starts anywhere with the first pickup on route and ends at depot (see Figure 7, XF and XFb lines in contrast to the complete X line).

Having the \( XF \) and \( XFb \) indicator variables bound by their open route quasi-balance constraints, we could streamline the restricted mixing constraints by imposing different capacity restrictions along route depending on the arc type (see Figure 7): For PP- and DP-links, we impose \( F[i,j] + Fb[i,j] \leq C * X[i,j] - \gamma * C * XF[i,j] \) and for DD-links \( F[i,j] + Fb[i,j] \leq C * X[i,j] - \gamma * C * XFb[i,j] \).

Results for up to 100 nodes using Gurobi (see Figure 8) show that all models solved by Wasan and Nagy 2014 are solved within 7.5% of optimizer’s time in the average (extremes are 50% of time for CE33_2 and 24 sec. instead of more than 8 hours for CE33_3). The equally optimal solutions are found with \( \tau = 5 \), thus a limited value of \( \tau \) suffices to obtain optimal and near optimal solutions with better quality. Models with 50 to 100 requests could be solved, as well. Restricted versions of VRPDP with a lower value of \( \tau \) are generally solved more efficiently. A three-in-one effect for lower \( \tau \) values is reached: efficient solution, better tour quality, and break of symmetry. A special LP fixing heuristic leading to nearly optimal solutions enables to solve larger models up to 200 nodes for VRPRMDP.

Merged network flows for VRPMDP (VRP with mixed deliveries and pickups) is the totality of the above presented techniques introducing indicator load variables, binding them along open paths, coupling them to existing flows, and efficiently encoding practical restrictions. Especially, the coupling of X, F, Fb, XF and XFb flows is what is meant by merging these heterogeneous flows. This stands in contrast to merging of homogeneous flows as we did by building Latest-First-Matches where many potential deadhead trip connections and flows are merged in one shared LFM-arc and in shared waiting arcs in connection lines of either connected stations. Thus “Merged network flows” is a unifying term for these kinds of flows’ merges which enable efficient solution of math models.
Appendix:

Figure 1: Trip-as-Node and Trip-as-Arc-Network for direct connections (at same station, here Halle)

Figure 2: The quadratic number of deadhead trips remains if only direct connections are merged

Figure 3: Merging deadhead trips by the LFM-Technique: First build First Matches, then Latest First Matches
Vehicle Scheduling with multiple Vehicle Types
- To each trip only some types are allowed

Multiple-Depot Vehicle Scheduling (MDVSP)
- Each Bus must return to the same depot of start

Solution Approach:
Multi-Commodity-Flow-Problem

Result:
Deadhead Aggregation by LFM network reduction
Flow models for practical schedule sizes directly solvable by standard Optimizers

Figure 4: Using LFM-Technique in Practice: mVT-MDVSP

Case of VRPB (VPR with Backhauling)
- Special case of VRPPD without Mixing

General Case VRPPD:
Introduce $\tau = \text{average number of DP-turns (LB-turns) per tour}$
Case: $\tau = 2$

Reducing $\tau$ makes the model solution faster in general

Figure 5: Visualizing the load operations: deliveries, then pickups for VRPB (Observation: 1 DP-turn)

Figure 6: Introducing the parameter $\tau$ of DP-turns to measure the quality of delivery and pickup routes
Figure 7: Efficient modeling of VRPDP and restricted Mixing using indicator variables XF and XFb, linked in one-side open routes respectively and coupled with the principle vehicle flow variables X.

Figure 8: Computational times for the direct solution of the Merged Flows Math Models for VRPDP using the Gurobi MIP Solver, computational results for up to 200 customers will also be presented.