An Inventory-Routing Problem with Pickups and Deliveries Arising in the Replenishment of Automated Teller Machines

Roel G. van Anholt
Faculty of Economics and Business Administration, VU University Amsterdam, 1081 HV Amsterdam, Netherlands,
roel.g.van.anholt@vu.nl

Leandro C. Coelho
CIRRELT and Faculté des sciences de l’administration, Université Laval, Québec, Québec G1V 0A6, Canada,
leandro.coelho@cirrelt.ca

Gilbert Laporte
CIRRELT and Canada Research Chair in Distribution Management, HEC Montréal, Montréal, Québec H3T 2A7, Canada,
gilbert.laporte@cirrelt.ca

Iris F. A. Vis
Faculty of Economics and Business, University of Groningen, 9747 AE Groningen, Netherlands,
iris.f.a.vis@rug.nl

The purpose of this paper is to introduce, model, and solve a rich multiperiod inventory-routing problem with pickups and deliveries motivated by the replenishment of automated teller machines in the Netherlands. Commodities can be brought to and from the depot, as well as being exchanged among customers to efficiently manage their inventory shortages and surpluses. A single customer can both provide and receive commodities at different periods, since its demand changes dynamically throughout the planning horizon and can be either positive or negative. In the case study, new technology provides these machines with the additional functionality of receiving deposits and reissuing banknotes to subsequent customers. We first formulate the problem as a very large-scale mixed-integer linear programming model. Given the size and complexity of the problem, we first decompose it into several more manageable subproblems by means of a clustering procedure, and we further simplify the subproblems by fixing some variables. The resulting subproblems are strengthened through the generation of valid inequalities and solved by branch and cut. We assess the performance of the proposed solution methodology through extensive computational experiments using real data. The results show that we are able to obtain good lower and upper bounds for this new and challenging practical problem.

Keywords: inventory-routing; vehicle routing; inventory management; pickup and delivery; branch and cut; clustering; exact algorithm; recirculation automated teller machines

History: Received: November 2013; revision received: March 2015; accepted: May 2015. Published online in Articles in Advance February 24, 2016.

1. Introduction

The purpose of this paper is to introduce, model, and solve an inventory-routing problem with pickups and deliveries (IRPPD) arising in the replenishment of automated teller machines (ATMs). Our study is motivated by the problem faced by a transporter responsible for such operations in the Netherlands. As in other cash-intensive economies, the Dutch credit institutions are gradually replacing regular ATMs by recirculation ATMs (RATMs), which are capable of accepting and dispensing banknotes, as well as checking their quality and authenticity. RATMs therefore provide customers the capability of both depositing and withdrawing cash. These machines provide tangible benefits to banks and customers, and are becoming the new standard in many markets. By the end of 2013, a total of 433,780 RATMs were in operation worldwide, corresponding to 17% of all ATMs (RBR 2014). The number of RATMs increased by 41% between 2012 and 2014, whereas the total number of ATMs grew by only 14% over the same period. RBR (2014) expects that the fast growth of the number and share of RATMs will continue to intensify. RATMs allow retailers to make deposits and enable customers to later withdraw the same cash. In this sense, RATMs are considerably more self-sufficient than the regular machines, which can only dispense cash. RATMs only require a visit to prevent the inventory level from reaching zero or from attaining the holding capacity of the machines when the supply and demand
are out of sync. When an RATM is empty, a customer can only use it to deposit cash, and when it is full only withdrawals are allowed.

In the problem considered here, the RATMs are replenished or partly emptied by using a fleet of rented armored trucks, which are very expensive. These trucks deliver cash from a depot to some machines, collect cash from some others to bring it back to the depot, or transfer cash between machines. The last operation reduces the routing cost and sometimes allows smaller or fewer trucks to be used. An important feature of the problem is the presence of inventory holding costs. Indeed, cash lying in a machine generates an implicit holding cost partly because it is insured and also because it incurs lost interest income.

To keep RATMs fully operational, that is when customers can use the RATM to both deposit and withdraw cash, one must solve an IRPPD. The IRPPD combines the features of two well-known classes of the vehicle routing problem: the inventory-routing problem (IRP) and the pickup and delivery problem (PDP), which we now briefly review.

The IRP belongs to the broader field of vendor-managed inventory systems in which a supplier coordinates the inventory management of a number of locations (Coelho, Cordeau, and Laporte 2014). In IRPs, the supplier must simultaneously decide when to visit its inventory locations, how much to deliver to each of them, and how to combine the deliveries into vehicle routes. There exist numerous variants of this problem (Andersson et al. 2010; Coelho, Cordeau, and Laporte 2014) and the related literature is rapidly expanding. The first exact algorithm for the single-vehicle IRP was based on branch and cut (Archetti et al. 2007). It could solve instances with up to 50 customers and three periods, and 30 customers and six periods. Archetti et al. (2012) later presented a hybrid tabu search matheuristic algorithm capable of dealing with larger instances and yielding solutions with very low optimality gaps. Coelho, Cordeau, and Laporte (2012) presented an adaptive large neighborhood search heuristic for the multivehicle IRP, and Coelho and Laporte (2013a) were the first to solve the multivehicle IRP exactly. Recently, multivehicle and multicombination IRPs were also solved to optimality by Coelho and Laporte (2013b). All of these algorithms are based on branch and cut. For a recent survey of models and algorithms, see Coelho, Cordeau, and Laporte (2014). A related problem appears in maritime transportation, see Christiansen, Fagerholt, and Ronen (2004) and Christiansen et al. (2013).

The PDP concerns the collection and distribution of one or several commodities to and from a set of locations. Berbeglia et al. (2007) classify PDPs and distinguish between three different problem structures: many-to-many (M-M), one-to-many-to-one (M-1-1), and one-to-one (1-1). The M-M structure means that each commodity may have multiple origins and multiple destinations, and that each location may be the origin or destination of multiple commodities. In 1-M-1 problems, some commodities are picked up at the depot and transported to some locations, whereas other commodities are picked up at these locations and transported to the depot. The 1-1 structure refers to a context in which each commodity has a single origin and a single destination, like in dial-a-ride problems (Cordeau and Laporte 2007).

Two important classes of PDPs with an M-M structure are the Swapping Problem (SP) and the One-Commodity Pickup and Delivery Traveling Salesman Problem (1-PDTSP). In the SP, introduced by Anily and Hassin (1992), each vertex of a graph provides a commodity and requests a commodity, possibly the same one. The problem is to design a least cost vehicle route to satisfy all requests. This problem is NP-hard on general graphs, but polynomial on some special structures (Anily, Gendreau, and Laporte 2011). Erdoğan, Laporte, and Cordeau (2010) have developed heuristics and a branch-and-cut algorithm for the multivehicle case. In the 1-PDTSP, each vertex either provides or requests a given amount of a single commodity, and a single vehicle route must be designed to transfer the right amounts of the commodity among vertices. The problem was introduced by Hernández-Pérez and Salazar-González (2003) and was solved by branch and cut (Hernández-Pérez and Salazar-González 2003, 2004a, 2007) and by heuristics (Hernández-Pérez and Salazar-González 2004b; Zhao et al. 2009). The 1-PDTSP arises in the rebalancing operations in shared bicycle systems (Benchimol et al. 2011; Chemla, Meunier, and Wölfle 2013; Contardo, Morency, and Rousseau 2012; Erdoğan et al. 2012; Erdoğan, Laporte, and Wölfle 2014; Raviv, Tzur, and Forma 2013). Some algorithms (Contardo, Morency, and Rousseau 2012; Dell’Amico et al. 2014; Raviv, Tzur, and Forma 2013; Shu et al. 2013) are capable of handling the multivehicle case. The problems encountered in RATM replenishment and bicycle rep positioning are similar in the sense that they both consist of moving a commodity between locations to keep its level within given limits. The main differences lie in inventory holding costs, which are larger in the case of RATMs, and also in the planning horizon. Indeed, in the case of RATMs, planning is typically done over several days whereas in the case of bicycles, repopositioning operations occur on a daily basis, sometimes only...
once during the night and often several times during the same day.

The PDP with a 1-M-1 structure deals with two types of commodities. Delivery commodities are transported from a single depot to multiple nodes, and pickup commodities are transported from multiple nodes back to the depot. A typical application is the recycling of products such as beer bottles, pallets, and containers (Berbeglia et al. 2007). A distinction is made between combined and single demands. Combined demands occur when locations may require both a pickup and a delivery, whereas single demand refers to problems where each location has either a pickup or delivery demand. The single demand problem was introduced by Mosheiov (1994). Heuristics were proposed by Gendreau, Laporte, and Vigo (1999) and by Subramanian and Battarra (2013), whereas Baldacci, Hadjiconstantinou, and Mingozzi (2003) and Hernández-Pérez and Salazar-González (2007) solved the problem exactly by branch and cut. The combined demand case was introduced by Min (1989). It was solved heuristically by Bianchessi and Righini (2007); Subramanian et al. (2010); Zachariadis, Tarantilis, and Kiranoudis (2010); Vidal et al. (2013), and exactly by branch-and-price algorithms by Angelelli and Mansini (2002); Dell’Amico, Righini, and Salani (2006); Subramanian and Battarra (2013). We are not aware of any contributions on the multimmodity 1-M-1-PDP.

Our aim is to model and solve the IRPPD arising in the replenishment of RATMs. We view it as a novel logistics application arising in the banking service sector. As will be seen, the Dutch application that motivates this study gives rise to a very large-scale mathematical model, which cannot be solved directly. To tackle it, we first decompose it into several subproblems through the application of a clustering procedure, and we further simplify the subproblems by fixing some variables. The resulting subproblems are then strengthened through the generation of valid inequalities and solved by branch and cut. The paper makes four main contributions. First, it introduces pickups and deliveries within an IRP context. Second, it combines two PDP structures: the 1-M-1 structure, which accounts for commodity movements from the depot to RATMs to the depot, and the M-M structure, which refers to commodity transfers among RATMs. Hence, our PDP could appropriately be designated as a 1-M-1-M problem. Note that we tackle a rather general case of this problem, in that there are several vehicles and side constraints in addition to the standard IRP and PDP features. Third, to cope with realistic sized instances, we propose a practical decomposition methodology combining clustering, variable fixing, and branch and cut. Our fourth contribution is to apply our algorithm to data derived from a real-world case arising in the Netherlands.

The remainder of this paper is organized as follows. We formally describe the problem in mathematical terms in §2, where we provide a mixed-integer formulation. The algorithm is described in §3. This is followed by the results of an extensive computational study in §4 and by conclusions in §5.

2. Mathematical Programming Formulation

The IRPPD is defined on a directed graph $G = (V, A)$, where $V$ is the set of vertices and $A$ the set of arcs. The set $V$ is partitioned into $\{D, R\}$, where $D$ is the set of depots and $R$ the set of RATMs. The set $A = \{(i, j): i, j \in V, i \neq j\}$ is the arc set. Each RATM incurs unit inventory holding costs $\alpha_t$ per period $(i \in R)$, and has an inventory holding capacity $C_i$. A handling cost $\beta_t$ is incurred at each depot, and is proportional to the quantity picked up and delivered at the depot. The length of the planning horizon is $p$, with discrete time periods $t \in T = \{1, \ldots, p\}$. A set of rented armored vehicles $k \in K = \{1, \ldots, K\}$, each with capacity $Q_k$ and average speed $s_k$, is given. A renting cost $\gamma_k$ per period is incurred if vehicle $k$ is used. Each vehicle is able to perform one route per period, from one depot to a subset of RATMs, each requiring $r$ units of time to be served, and back to the same depot. The shift of each vehicle is limited to 5 minutes, after which monetary units per minute of overtime are incurred. A travel time $c_{ij}$ in minutes is associated with arc $(i, j) \in A$. We assume each depot has sufficient inventory and capacity to perform all pickups and deliveries during the planning horizon. The inventories are not allowed to exceed the holding capacity nor are they allowed to become negative. At the beginning of the planning horizon, the decision maker knows the current inventory level $I_0$ of the RATMs and receives information on the net demand $d'_i$ of each RATM $i$ for each period $t$. Negative demands mean that the RATM provides a commodity, whereas positive demands mean that the RATM receives some quantity of the commodity. We assume that the quantities $q'_i$ received by RATM $i$ in period $t$ can be used to satisfy its net demand in that period. The quantities picked up at the depots and at the RATMs may be delivered to any RATM to satisfy their demands. The objective of the problem is to minimize the total cost while satisfying the net demand for each RATM in each period.

The variables used in the formulation are as follows. Five families of binary variables are used: directed routing variables $x_{ij}^k$ are equal to 1 if and only if arc $(i, j)$ is used on the route of vehicle $k$ in period $t$; visiting variables $y_i^t$ are equal to 1 if and only if RATM $i$ is visited by vehicle $k$ in period $t$; vehicle usage variables $z^k_t$ are equal to 1 if and only if vehicle $k$ is used in period $t$; delivery variables $w_{ij}^k$ are equal to 1 if and only if a delivery is made to RATM $i$ by vehicle $k$ in
period \( t \); and pickup variables \( z_{kj}^{kt} \) are 1 if and only if a pickup is performed at RATM \( i \) by vehicle \( k \) in period \( t \).

Integer variables \( l_i^{kt} \) represent the inventory level at RATM \( i \in \mathcal{R} \) at the end of period \( t \in \mathcal{T} \). Variables \( q_{kl}^{kt} \) are integers representing the product quantity delivered to RATM \( i \) using vehicle \( k \) in period \( t \), and variables \( p_{kl}^{kt} \) are integers representing the product quantity picked up from RATM \( i \) using vehicle \( k \) in period \( t \). Two sets of variables represent the amount of inventory carried by vehicle \( k \) in period \( t \) into and out of the depot: \( H_i^{kt} \) and \( J_i^{kt} \), respectively. The load of vehicle \( k \) after serving RATM \( i \) in period \( t \) is represented by variables \( u_{ij}^{kt} \), and \( E_k^t \) represents the overtime in number of extra minutes worked by vehicle \( k \) in period \( t \) beyond the regulatory shift duration \( S \).

The mult depot IRPPD is then formulated as follows:

\[
\text{(MD-IRPPD)}
\]

minimize \( \sum_{i \in \mathcal{S}} \sum_{t \in \mathcal{T}} \alpha_i l_i^{t-1} + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \beta(l_i^{kt} + H_i^{kt})^2 \]

\[
+ \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \gamma_k v^{kt} + \sum_{k \in \mathcal{K}} \delta_k E_k^t \}
\]

subject to

\[
l_i^{kt} = l_i^{t-1} + \sum_{j \in \mathcal{R}} q_{ij}^{kt} - \sum_{j \in \mathcal{R}} p_{jik}^{kt} + d_i, \quad i \in \mathcal{R}, \quad t \in \mathcal{T},
\]

\[
0 \leq l_i^{kt} \leq C_i, \quad i \in \mathcal{R}, \quad t \in \mathcal{T},
\]

\[
\sum_{j \in \mathcal{R}} q_{ij}^{kt} - \sum_{j \in \mathcal{R}} p_{jik}^{kt} = 0, \quad i \in \mathcal{V}, \quad k \in \mathcal{K}, \quad t \in \mathcal{T},
\]

\[
\sum_{j \in \mathcal{V}} x_{ij}^{kt} = y_i^{kt}, \quad i \in \mathcal{V}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
\sum_{i \in \mathcal{R}} x_{ij}^{kt} \leq 1, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
y_i^{kt} \leq u_{ij}^{kt}, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
u_i^{kt} \leq \sum_{i \in \mathcal{R}} x_{ij}^{kt}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
\sum_{i \in \mathcal{R}} x_{ij}^{kt} \leq \sum_{i \in \mathcal{R}} y_i^{kt} - y_m, \quad \mathcal{F} \subseteq \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T}, \quad m \in \mathcal{T},
\]

\[
w_{ij}^{kt} \leq y_i^{kt}, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
z_i^{kt} \leq y_i^{kt}, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
q_{ij}^{kt} \leq w_{ij}^{kt}(C_i - l_i^{kt}), \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
p_{ij}^{kt} \leq z_{ij}^{kt} l_i^{kt}, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
w_{ij}^{kt} + z_{ij}^{kt} \leq 1, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
s_k \sum_{(i,j) \in \mathcal{E}} c_{ij} x_{ij}^{kt} + r \sum_{j \in \mathcal{R}} y_{ij}^{kt} \leq S + E_k^t, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
u_i^{kt} \geq (u_i^{kt} + p_i^{kt} - q_i^{kt}) x_{ij}^{kt}, \quad i, j \in \mathcal{V}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
0 \leq u_{ij}^{kt} \leq Q_k, \quad i \in \mathcal{V}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
J_k^t = \sum_{i \in \mathcal{S}} u_{ij}^{kt}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
H_k^t = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{S}} x_{ij}^{kt} u_{ij}^{kt}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
q_{ij}^{kt} - p_{ij}^{kt} + H_i^t \geq 0, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
x_{ij}^{kt} \in \{0, 1\}, \quad (i, j) \in \mathcal{E}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
u_i^{kt}, y_i^{kt} \in \{0, 1\}, \quad i \in \mathcal{V}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T},
\]

\[
\omega_i^{kt}, z_i^{kt} \in \{0, 1\}, \quad i \in \mathcal{R}, \quad k \in \mathcal{H}, \quad t \in \mathcal{T}.
\]

The objective function (1) minimizes the total cost of inventory holding, inventory handling at the depot, and vehicle renting. Constraints (2) state the inventory conservation condition over successive periods: they define the inventory in period \( t \) as the inventory held in period \( t - 1 \), plus the quantity delivered, minus the quantity picked up, plus the net demand of the location. Constraints (3) define the bounds on the inventory held by each RATM throughout all periods. Constraints (4)–(9) guarantee that proper vehicle routes are created. More specifically, constraints (4) are flow conservation constraints, constraints (5) relate the incoming flow of a vertex to the fact that it is visited or not. Constraints (6) state that a vehicle can leave at most one depot in any period. Constraints (7) mean that a vertex cannot be visited if no vehicle is assigned to it. Similarly, constraints (8) relate the use of a vehicle during a given period to the flow variables associated with that vehicle and that period. Constraints (9) prevent the formation of subtours over a set of RATMs. These are the subtour elimination constraints proposed by Gendreau, Laporte, and Semet (1997) for the covering tour problem. Constraints (10) link the delivery binary variables \( w_{ij}^{kt} \) to the visiting binary variables \( y_i^{kt} \). They allow a delivery decision to be made only if a visit is performed to the RATM. Constraints (11) are similar to (10) and apply to the pickup decisions. Constraints (12) and (13) allow a quantity to be delivered or picked up at an RATM only if the corresponding binary decision variable is equal to 1. Constraints (14) ensure that a visit to an RATM is made either for a pickup or for a delivery operation. Constraints (15) guarantee that the shift duration is respected and that overtime is properly accounted for. Constraints (16) ensure that the load within the vehicle is consistent along its route. Constraints (17) set bounds on the load inside the vehicle after serving RATM \( i \). Constraints (18) set the load of the vehicle when leaving the depot, whereas constraints (19) compute its load when returning to the depot at the end of the route. Finally, constraints (20)–(23) define nonnegativity and binary conditions on the variables.

### 3. Solution Algorithm

In the RATM application that motivates this study, 6,377 machines must be replenished by 32 vehicles,
each based at a different depot, over a six-day planning horizon. This means that the (MD-IRPPD) model just described will contain more than 7.8 billion binary variables and about the same number of nonlinear constraints, in addition to $O(2^n)$ subtour elimination constraints. Solving such a model or even its continuous relaxation is beyond the reach of any available mathematical programming solver. We have therefore opted for a practical three-step solution strategy. We first apply a clustering procedure to decompose the initial problem into 32 subproblems of about 200 RATMs each, one for each depot (each having a single vehicle). This reduces the number of variables of each subproblem to a more manageable size, around 250,000, but still too many to reach optimality. For comparison purposes, the largest IRP instance that can be solved exactly, with a single vehicle and a single depot, contains 200 customers (Coelho and Laporte 2014), but the IRP does not contain many of the features presented in this paper. Likewise, the largest M-M PDP instance solved to optimality contains 200 vertices and a single vehicle (Hernández-Pérez and Salazar-González 2007). The IRPPD combines features of these two problems and is significantly more complicated than either of these. For these reasons, it is unrealistic to solve this problem exactly for the large sizes we are considering. We therefore simplify the subproblems further by fixing some variables. This yields instances that can be solved exactly or to near optimality by branch and cut. These three phases are described in §§3.1–3.3, respectively. Algorithm 1 provides a sketch of our solution procedure.

Algorithm 1 (Overall solution algorithm)

1: Clustering procedure: solve a transportation problem to obtain as many single-depot instances as the number of depots (see §3.1)
2: for each single-depot instance do
3: Apply the variable fixing procedure (see §3.2) to reduce the instance size
4: Solve the IRPPD model by branch and cut with some fixed variables and valid inequalities (see §3.3)
5: end for.

3.1. Clustering Procedure

Partitioning the customer base into clusters not only makes sense from a computational point of view but also from an empirical standpoint because this is how the problem is solved in practice. We solve a transportation problem to assign RATMs to depots while controlling the size of the clusters. The assignment cost in each cluster is the sum of the radial average travel times, interpreted here as distances. Eilon, Watson-Gandy, and Christofides (1971) show that radial distances are often a good proxy for routing costs. Let $\pi_{ij}$ be a binary variable equal to 1 if and only if RATM $i$ is assigned to depot $j$. The transportation problem is then

\[
\begin{align*}
\text{minimize} & \quad \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{D}} (c_{ij} + c_{ij}) \pi_{ij}/2 \quad (24) \\
\text{subject to} & \quad \sum_{j \in \mathcal{D}} \pi_{ij} = 1, \quad i \in \mathcal{R}, \quad (25) \\
& \quad \sum_{i \in \mathcal{R}} \pi_{ij} \leq A, \quad j \in \mathcal{D}, \quad (26) \\
& \quad \pi_{ij} \in [0,1], \quad i \in \mathcal{R}, j \in \mathcal{D}. \quad (27)
\end{align*}
\]

The objective function (24) minimizes the total sum of the distances between assigned RATMs and depots. Constraints (25) ensure that each RATM is assigned to exactly one depot, and constraints (26) ensure that each depot is assigned to at most $A$ RATMs, with $A$ chosen so as to balance the number of RATMs per depot. Constraints (27) define the domain of the variables.

After the clustering procedure has been applied, the problem is simplified and the (MD-IRPPD) formulation can be reduced to one without the depot index, i.e., $\mathcal{D} = \{0\}$. We provide this formulation for the sake of completeness and because we will later use it to generate valid inequalities. Reusing the notation employed in §2, we now state the single-depot formulation (IRPPD)

\[
\begin{align*}
\text{minimize} & \quad \left\{ \sum_{i \in \mathcal{R}} \sum_{t \in \mathcal{T}} \alpha_i l_i^t + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \beta(l_k^t + H_k^t) + \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} \gamma_i y_i^k + \sum_{k \in \mathcal{K}} \delta E_k \right\} \quad (28) \\
\text{subject to} & \quad l_i^t = l_i^{t-1} + \sum_{k \in \mathcal{K}} q_i^k - \sum_{j \in \mathcal{V}} p_i^{k,t} + d_i^t, \quad i \in \mathcal{R}, t \in \mathcal{T}, \quad (29) \\
& \quad 0 \leq l_i^t \leq C_i, \quad i \in \mathcal{R}, t \in \mathcal{T}, \quad (30) \\
& \quad \sum_{j \in \mathcal{V}} x_i^k(t) - \sum_{j \in \mathcal{V}} x_{ij}^k = 0, \quad i \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (31) \\
& \quad \sum_{j \in \mathcal{V}} x_{ij}^k = y_i^k, \quad i \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (32) \\
& \quad \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{V}} x_{ij}^k \leq \sum_{t \in \mathcal{T}} \sum_{j \in \mathcal{V}} y_{ij}^k - y_{ij}^m, \quad \mathcal{F} \subseteq \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (33) \\
& \quad w_{ij}^k \leq y_{ij}^k, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (34) \\
& \quad z_i^k \leq y_i^k, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (35) \\
& \quad q_i^k \leq w_{ij}^k (C_i - l_i^t), \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (36) \\
& \quad p_i^{k,t} \leq z_i^k H_i^k, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (37) \\
& \quad w_{ij}^k + z_i^k \leq 1, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (38) \\
& \quad s_k \sum_{(i,j) \in \mathcal{E}} \ell_{ij}^k + r \sum_{i \in \mathcal{R}} y_{ij}^k \leq S + E_k, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad (39) \\
& \quad u_j^k \geq (u_j^k + p_i^{k,t} - q_i^k) x_{ij}^k, \quad i, j \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (40)
\end{align*}
\]
0 \leq u_{ij}^{kt} \leq Q_k, \quad i \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (41)
\sum_{i \in \mathcal{K}} u_{ij}^{kt} = 1, \quad j \in \mathcal{J}, \quad (42)
H_i^t = \sum_{k \in \mathcal{K}} x_{ij}^{kt}, \quad k \in \mathcal{K}, t \in \mathcal{T}, \quad (43)
q_i^{kt}, p_i^{kt}, j_k^t, H_i^t \geq 0, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (44)
x_{ij}^{kt} \in \{0, 1\}, \quad (i, j) \in \mathcal{A}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (45)
y_{ij}^{kt} \in \{0, 1\}, \quad i \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (46)
w_{ij}^{kt}, z_{ij}^{kt} \in \{0, 1\}, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (47)

The description of the variables, objective function, and constraints is similar to that of the multidepot model. Three comments are relevant with respect to the IRPPD formulation. The first regards constraints (40), which resemble the Miller-Tucker-Zemlin (1960) subtour elimination constraints. However, these can be linearized as follows. Constraints (36) and (37) can be rewritten in a slightly weaker form as
\begin{align*}
q_i^{kt} &\leq w_i^{kt}c_i, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (48) \\
p_i^{kt} &\leq z_i^{kt}c_i, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (49)
\end{align*}

On their own, these constraints would allow quantities to be picked up or delivered to violate the inventory bounds, but together with constraints (29) and (30), they are feasible linear representations of constraints (36) and (37). Constraints (40) can be linearized as
\begin{align*}
u_{ij}^{kt} &\geq u_{ij}^{kt} + p_{ij}^{kt} - q_{ij}^{kt} - (1 - x_{ij}^{kt})Q_k, \\
&\quad i, j \in \mathcal{V}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (50)
\end{align*}

Finally, constraints (43) can be rewritten in a linear form as
\begin{align*}
H_i^t &\geq u_{ij}^{kt} - (1 - x_{ij}^{kt})Q_k, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (51) \\
H_i^t &\leq Q_k, \quad k \in \mathcal{K}, t \in \mathcal{T}. \quad (52)
\end{align*}

The IRPPD formulation can be strengthened through the generation of the following valid inequalities:
\begin{align*}
x_{ij}^{kt} &\leq y_{ij}^{kt}, \quad i, j \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}, \quad (53) \\
y_{ij}^{kt} &\leq y_0^{kt}, \quad i \in \mathcal{R}, k \in \mathcal{K}, t \in \mathcal{T}. \quad (54)
\end{align*}

Constraints (53) are referred to as logical inequalities. They tighten the relation between routing and visiting variables. Constraints (54) include the supplier in the route of vehicle k if any RATM is visited by that vehicle in that period.

### 3.2. Variable Fixing Procedure

To solve realistic instances, we have designed a variable fixing procedure, which is applied prior to executing the branch-and-cut algorithm, in the spirit of granular search (Johnson and McGeoch 1997; Toth and Vigo 2003). This heuristic determines for each RATM \( i \in \mathcal{R} \) and each period \( t \in \mathcal{T} \) whether it must be visited, must not be visited, or must perhaps be visited. Specifically, for each period \( t \) we distinguish five subsets: (1) a must pickup set \( \mathcal{U}^t \), (2) a must deliver set \( \mathcal{W}^t \), (3) a perhaps pickup set \( \mathcal{U}^t \), (4) a perhaps deliver set \( \mathcal{W}^t \), and (5) a not visit set \( \mathcal{L}^t \). These five sets define a partition of \( \mathcal{R} \). When RATMs are added to one of the subsets \( \mathcal{U}^t, \mathcal{W}^t, \) or \( \mathcal{L}^t \), the size of the formulation is considerably reduced and the branch-and-cut algorithm benefits from this reduction, because some of the decisions related to pickups, deliveries, and visits for some RATMs are already made. When RATMs are added to one of the other two subsets \( \mathcal{U}^t \) or \( \mathcal{W}^t \), the algorithm still determines for each period \( t \) whether a visit is required, but the benefit of the variable fixing procedure is that it limits the choice to either a pickup or a delivery. It is implemented as follows:
\begin{align*}
\sum_{k \in \mathcal{K}} z_{ij}^{kt} &= 1, \quad i \in \mathcal{U}^t, t \in \mathcal{T}, \quad (55) \\
\sum_{k \in \mathcal{K}} w_{ij}^{kt} &= 1, \quad i \in \mathcal{W}^t, t \in \mathcal{T}, \quad (56) \\
\sum_{k \in \mathcal{K}} w_{ij}^{kt} &= 0, \quad i \in \mathcal{U}^t, t \in \mathcal{T}, \quad (57) \\
\sum_{k \in \mathcal{K}} z_{ij}^{kt} &= 0, \quad i \in \mathcal{W}^t, t \in \mathcal{T}, \quad (58) \\
\sum_{k \in \mathcal{K}} y_{ij}^{kt} &= 0, \quad i \in \mathcal{L}^t, t \in \mathcal{T}. \quad (59)
\end{align*}

Also, it follows directly that several other constraints fixing binary variables can be derived from (55)–(59). These are used to further reduce the size of the problem by effectively eliminating several variables from the problem
\begin{align*}
\sum_{k \in \mathcal{K}} y_{ij}^{kt} &= 1, \quad i \in \mathcal{U}^t, t \in \mathcal{T}, \quad (60) \\
w_{ij}^{kt} &= 0, \quad i \in \mathcal{W}^t, k \in \mathcal{K}, t \in \mathcal{T}, \quad (61) \\
\sum_{k \in \mathcal{K}} y_{ij}^{kt} &= 1, \quad i \in \mathcal{W}^t, t \in \mathcal{T}, \quad (62) \\
z_{ij}^{kt} &= 0, \quad i \in \mathcal{W}^t, k \in \mathcal{K}, t \in \mathcal{T}, \quad (63) \\
w_{ij}^{kt} &= 0, \quad i \in \mathcal{L}^t, k \in \mathcal{K}, t \in \mathcal{T}. \quad (64)
\end{align*}
The first parameter relates to due periods and tipping periods. A due period is a period in which a visit is required to prevent the inventory from exceeding the holding capacity or from a stock out. A tipping period is a period in which a pickup is required to remain cost efficient, that is when the holding cost is disproportionately high related to the cost of visiting the RATM. The parameters are as follows:

- $m$: number of periods preceding a due period or a tipping period;
- $f$: number of RATMs that are closest to RATM $i$;
- $g$: minimum inventory level of RATM $i$, in a percentage of its holding capacity $C_i$;
- $b$: expected number of RATM visits per vehicle $k$ in a period $t$.

In what follows, let $I_i^t = I_i^0 + \sum_{j=t}^t d_i^j$ be the cumulative inventory of RATM $i$ in period $t$. This is useful to identify up to which period $t$ the RATM will be capable of respecting its inventory constraints without replenishment. The pseudocode of the variable fixing procedure is presented in Algorithm 2.

Algorithm 2 (Variable fixing procedure ($m$, $f$, $g$, $b$))

1:  \( \mathcal{R} \leftarrow \emptyset \)
2:  \textbf{for all } $i \in \mathcal{R}$ \textbf{do}
3:     \textbf{for all } $t \in \mathcal{T}$ \textbf{do}
4:         \textbf{Add } $i$ \textbf{to } $\mathcal{X}^t$
5:         \textbf{if } $t = 1$ \textbf{or } $m = 0$ \textbf{then}
6:             \textbf{if } $I_i^t > C_i$ \textbf{then}
7:                 \textbf{Add } $i$ \textbf{to } $\mathcal{W}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
8:             \textbf{else if } $I_i^t < 0$ \textbf{then}
9:                 \textbf{Add } $i$ \textbf{to } $\mathcal{W}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
10: \textbf{end if}
11: \textbf{else}
12:     \textbf{for } $t' = t - m$ \textbf{to } $t' = t$; $t' > 1$ \textbf{do}
13:         \textbf{if } $\sum_{e \in \mathcal{E}} d_i^{t'} > 0 \text{ and } (I_i^0 \times a_i) > (g_i/b)(I_i^0/((\sum_{e \in \mathcal{E}} d_i^{t'})/p))$ \textbf{then}
14:             \textbf{Add } $i$ \textbf{to } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
15:         \textbf{else if } $I_i^t > C_i$ \textbf{then}
16:             \textbf{Add } $i$ \textbf{to } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
17:         \textbf{else if } $I_i^t < 0$ \textbf{then}
18:             \textbf{Add } $i$ \textbf{to } $\mathcal{Y}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
19: \textbf{end if}
20: \textbf{end for}
21: \textbf{end if}
22: \textbf{for all } $f$ \textbf{number of RATMs } $j$ \textbf{having the smallest } $c_j$ \textbf{to } $i$ \textbf{do}
23:     \textbf{if } $i \in \{\mathcal{W}^t, \mathcal{X}^t\}$ \textbf{and } $(I_i^0/C_i) < (1-g)$ \textbf{then}
24:         \textbf{Add } $j$ \textbf{to } $\mathcal{Y}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
25:     \textbf{else if } $i \in \{\mathcal{W}^t, \mathcal{Y}^t\}$ \textbf{and } $(I_i^0/C_i) > g$ \textbf{then}
26:         \textbf{Add } $j$ \textbf{to } $\mathcal{Y}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{X}^t$; \textbf{remove } $i$ \textbf{from } $\mathcal{R}$
27: \textbf{end if}
28: \textbf{end for}
29: \textbf{end for}
30: \textbf{end for}
31: \textbf{Return } $\mathcal{W}^t, \mathcal{W}^t, \mathcal{X}^t, \mathcal{Y}^t, \mathcal{Z}^t$.

The variable fixing procedure starts by adding each RATMs $i$ in each period $t$ to the not visit subset $\mathcal{X}^t$ (lines 2–4). When $t = 1$ or $m = 0$, lines 5–10 ensure that RATMs are added to the must visit subsets $\mathcal{W}^t$ and $\mathcal{W}^t$ when the holding capacity $C_i$ is exceeded or when the inventory $I_i^t$ becomes negative. If either condition is satisfied, the algorithm does not have the flexibility to choose between preceding or succeeding periods to visit the RATM. Therefore, the RATM must be visited.

If the condition in line 5 is not met, we check three more conditions in lines 13–18 to add RATMs $i$ to the perhaps visit subsets $\mathcal{X}^t$ and $\mathcal{Y}^t$ in period $t$ as well as to the subsets in $m$ number of preceding periods $t$. Line 12 ensures that RATMs $i$ are also added to the subsets in $m$ number of preceding periods $t$. We introduce lines 13 and 14 to ensure RATM $i$ is visited for a pickup when a tipping period occurs, i.e., when the holding cost exceeds the cost of visiting the RATM. More precisely, the following condition is verified: the inventory $I_i^t$ of RATM $i$ in period $t$ multiplied with the unit inventory holding cost $a_i$ exceeds the vehicle rental cost $g_i$ divided by the expected number of visits per route $b$ and divided by the expected number of days elapsed since the last visit. To this end, we calculate the expected number of days elapsed since the last visit by dividing the inventory level $I_i^t$ by the average demand $\sum_{e \in \mathcal{E}} d_i^{t'}/t$. If the latter condition is met, then the RATM $i$ is added to the perhaps pickup set $\mathcal{X}^t$. RATMs $i$ are also added to the $\mathcal{X}^t$ subset in periods $t$ when the inventory exceeds capacity $I_i^t > C_i$ (lines 15 and 16). In the case of stock outs, the RATMs are added to the perhaps deliver subset $\mathcal{Y}^t$, which is indicated in lines 17 and 18.

Most likely, some RATMs $i$ will have been taken out of set $\mathcal{X}^t$ and added to the other subsets $\mathcal{W}^t, \mathcal{W}^t, \mathcal{X}^t$, and $\mathcal{Y}^t$ when arriving at line 18. The final part in lines 22–28 ensures that even more RATMs $j$, which are located closest to RATMs $i$ belonging to $\mathcal{W}^t, \mathcal{W}^t, \mathcal{X}^t$, or $\mathcal{Y}^t$, are added to the perhaps visit subsets. The number of RATMs $j$ added per RATM $i$ is determined...
by parameter $f$. It should be noted that RATMs $j$ are only added when they meet an inventory level $I_{ij}$, which is at least as high as a given percentage $g$ of the holding capacity $C_i$ for a pickup, or lower than a given percentage $1 - g$ for a delivery. The rationale of this final part is to stimulate the exchange of inventories between RATMs $i$ and $j$ to eventually reduce the amount to be picked up from, and delivered to the depot.

3.3. Branch-and-Cut Algorithm

We have implemented a branch-and-cut algorithm capable of solving the IRPPD model with the linearized constraints, the valid inequalities, and the fixed variables. All variables of the formulation are explicitly handled by the algorithm, but we cannot generate all subtour elimination constraints (33) a priori. These will be dynamically generated as cuts as they are found to be violated. The formulation is then solved by branch and cut as follows. At a generic node of the search tree, a linear program with relaxed integrality constraints is solved, a search for violated constraints is performed, and violated valid inequalities are added to the current program, which is then reoptimized. This process is reiterated until a feasible or dominated solution has been reached, or until no more cuts can be added. At this point, branching on a fractional variable occurs. Details regarding the implementation and improvements to this algorithm are provided in §4.2.

4. Computational Experiments

We will present the instances generator in §4.1, some implementation details in §4.2, the analysis of our extensive computational experiments in §4.3, and the estimated benefits of our study in §4.4. The instance set as well as detailed computational results are available at http://www.leandro-coelho.com/instances.

4.1. Instances Description

The data used in our experiments stem from a real-world case in the Netherlands. A total of 6,377 cash dispensing self-service devices, both regular and recirculation ATMs, were installed in the Netherlands in 2013 (ABN ARMO Bank 2013; ING Bank 2013; Rabobank 2013). These are replenished from 32 cash centers (depots). By setting $A$ equal to 200 in constraints (26), we assigned the 6,377 RATMs fairly evenly to the depots, each of which serves a small region of the Netherlands (Figure 1). For our experiments, we assume that all self-service devices in the Netherlands are RATMs. This is not unrealistic given the recent strong increase in the share of RATMs in the Netherlands (ECB 2013). Also, in several other countries, such as Japan, RATMs already dominate the cash self-service device market (RBR 2014). To conduct our numerical experiments, we used real data for the locations of RATM devices, true distances and speeds, and real RATM demands. Although the Netherlands is rather densely populated, not all parts of it are equally well served by RATMs. Their coverage varies considerably over the country. Our solution methodology is rather robust and can easily deal with this diversity. Figure 2 depicts the position of the RATMs and the cash center for the Amsterdam area. There still exist some ARMs that are not RATMs. For these, we generated RATM demands. The depot locations are confidential for security reasons, so we used approximate locations.

To mimic real demands for devices that are not yet RATMs, we have carefully studied real demands from readily installed RATMs. This study demonstrates that the number of transactions per day is well estimated by a Poisson probability distribution; the number of deposits and the number of withdrawals follow completely different patterns in terms of quantity, value, and weekly seasonality pattern; a huge diversity exists among the demand intensity of RATMs, and the average value per deposit is €1,000 and the average value per withdrawal is €133.

Based on these findings, we have generated deposit and withdrawal distributions separately and added these sequentially to construct a net demand per RATM per day. For the deposits and withdrawals of each RATM, we uniformly assigned two demand intensity levels of one to five to mimic demand intensity for both deposits and withdrawals (Table 1).

By randomly drawing from a Poisson distribution with the assigned number of transactions per day and multiplying the resulting number of transactions by the average transaction amount, i.e., €1,000 or €133, we obtained daily deposit and withdrawal amounts. Finally, to cope with week seasonality, we modified these amounts with the factors described in Table 2.

The parameters of our instance are then as follows:

- $\alpha_i = €0.08$ per €1,000 inventory per period $t$ (based on a 3% annual interest rate).
- $C_i = €260,000$ per RATM $i$. This is an estimation based on an RATM with four cassettes, each with a capacity of 2,000 notes.
- $\beta = €0.30$ per €1,000 picked up or delivered at the depot.
- $p = $six periods. Periods coincide with days, which together define a workweek from Monday to Saturday. In practice, order lead times are not more than one or two days and so a planning horizon of six days is sufficient.
- $K = $one vehicle.
- $Q_k = €7,800,000$ per vehicle $k$. This is an estimation based on 30 full replenishments of €260,000 in a single period.
• $\gamma_k = \€2,000$ per vehicle $k$ per period used.
• $r = 18$ minutes.
• $S =$ eight hours. This is the regular daily work time in the Netherlands. We assume the vehicle is loaded prior to performing the route, so the driver has at most eight hours to perform all pickups and deliveries.
• $\delta = \€8.00$ per minute, which is approximately twice as expensive as the regular vehicle renting cost.
• $I^0_i \in \{0, \ldots, C_i\}$. Each RATM $i$ is assigned a random initial inventory in euros.
• $d_j^p = \{\pm 45,000, \ldots, \pm 45,000\}$ per period. Random values are drawn from a Poisson distribution for both
withdrawals and deposits, which are thereafter combined into a net demand. To simulate real demands at different locations, we have ensured that RATMs either have a net-positive, net-negative, or balanced demand without losing stochasticity. We have then generated different demands at distinct periods: in periods $t \in \{1, 2\}$ the demand tends to be net-positive, because in the Netherlands more cash is deposited on Mondays and Tuesdays. The demand in periods $t \in \{5, 6\}$ tends to be net-negative, since more cash is withdrawn on Fridays and Saturdays.

- $c_{ij}$ = the arc set is constructed by first using real travel distances between the RATMs and the depot. A routable network data set for the Netherlands was then constructed using OpenStreetMap data (Geofabrik GmbH, OpenStreetMap Contributors 2013) and average driving speeds were used on the various road types to approximate true speeds.

### 4.2. Implementation Features

The algorithm just described was coded in C++ using IBM Concert Technology and solved with CPLEX 12.5.1 running on a single thread. All computations were

<table>
<thead>
<tr>
<th>Demand intensity level</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deposits per day</td>
<td>10.0</td>
<td>17.5</td>
<td>25.0</td>
<td>32.5</td>
<td>40.0</td>
</tr>
<tr>
<td>Withdrawals per day</td>
<td>75.0</td>
<td>131.3</td>
<td>187.5</td>
<td>243.8</td>
<td>300.0</td>
</tr>
</tbody>
</table>
executed on a grid of Intel Xeon processors running at 2.66 GHz with up to 48 GB RAM installed per node, with the Scientific Linux 6.1 operating system. A time limit of six hours was imposed on the execution of each of the 32 instances.

The clustering procedure was run only once. It is conceivable that overall improvements could have been obtained by performing 1-opt or 2-opt moves between clusters, but there exist \( O(|X|^2) \) such potential moves and evaluating each of them would have entailed solving a large-scale integer linear program, which was impractical.

Several experiments were performed with various variable fixing parameters. These settings were gathered from cash supply chain parties in the Netherlands, and estimated when these could not be made publicly available for security reasons. The parameters for the clustering heuristic were set as follows: \( m \in \{0, 1, 2\}; f \in \{0, 1, 2\}; g \in \{30, 50, 70\}; b \in \{15\} \).

For the instances considered in this paper, which contain up to 200 vertices, we cannot obtain optimal solutions with our branch-and-cut algorithm. To cope with this situation, we have decided to add two new layers to the branch-and-cut algorithm. In the first one, we verify at every node having a fractional solution whether inequalities (53) are violated and we then add them to the linear program. This helps improve the lower bound of the instance. The second new layer is added to handle constraints (50) since we have observed that these constraints are not tight with respect to the vehicle capacity and are generally satisfied. For this reason, we have devised a procedure to verify whether they are violated only at an integer solution. If the integer solution is found to violate constraints (50), these are added and the node is then reoptimized. Otherwise, the solution is feasible for the IRPPD. These two enhancements have significantly decreased the computational time.

Moreover, we have also observed that the separation algorithm of constraints (9) can be improved by adding a new set of variables representing the route of the vehicle in an undirected graph. By doing this, the separation algorithm runs on a graph half the size of the original one, looking for connected components and deriving maximum cuts over a much smaller network. Moreover, each subtour elimination cut derived for this new variable is equivalent to two cuts expressed in the original variables, one in each direction.

These two procedures employed to generate new cuts dynamically typically yield around 20,000 cuts only at the root node of the search tree. This number, although sizeable, is only a small fraction of the total number of cuts that could potentially be generated. By optimizing a problem with fewer constraints, we gain in speed since far fewer simplex iterations are needed. We note that a typical IRPPD instance still contains around 250,000 binary variables and 150,000 constraints after the two procedures just described have been applied, besides \( O(2^{|X|}) \) subtour elimination constraints.

We have also observed that at the beginning of the optimization process, the problem is degenerate, i.e., several pivoting operations do not improve the value of the objective function. We have therefore taken advantage of the fact that during the optimization process, one can change the routing decisions while remaining feasible. Since routing variables \( x \) do not appear in the objective function, this solution has the same objective function value. Note that as long as the total route length is less than the shift duration, the solution remains feasible. We have also tested adding a small coefficient to the objective function to further minimize route length, and thus avoid degeneracy, but this option did not yield any significant advantage.

Finally, to obtain a clear view of the performance of our algorithms, and also to speed up the solution of each node, we have turned off the CPLEX cut generation.

4.3. Analysis of the Computational Experiments

We start our analysis by presenting the results obtained by the algorithm presented in §§3 and 4.2 on the set of 32 instances described in §4.1. Not all conceivable instances of the MD-IRPPD are feasible, but feasibility was always achieved with our data. The optimality gaps observed before variable fixing are rather large, with an average of 51%, a maximum of 62%, and a minimum of 45%. Even when 24 hours of computing time were allotted, these figures did not improve significantly: the average lower bound increased by 2% from 16,002 to 16,355, and the average gap went down from 51.31% to 50.23%. In what follows, we use the information obtained within six hours of computing time to allow for a direct comparison with the performance of the variable fixing procedure. To benchmark the quality of these solutions, we can compare to the instances of the IRP without pickup and delivery, which has recently been solved to optimality in Coelho and Laporte (2014) for comparable sizes. The problem at hand is much more complicated and clearly requires an additional effort to obtain good solutions.

We have then limited the flexibility of the algorithm by disallowing visits to RATMs that do not require a pickup or delivery to remain operational. This is achieved by setting to zero both parameters \( m \) and \( f \) of the clustering procedure. Obviously, all solutions
obtained for this constrained version of the problem remain valid for the general case. In this situation, we have observed an average reduction of 24\% in the upper bounds, even if the problem is more constrained. Note that this improvement was not observed even when the computing time was four times as long for the case without variable fixing. This can be explained by the fact that the branch-and-cut algorithm performs a search in a more promising area of the solution space. Under this scenario, seven instances were solved to optimality, and most of them yielded gaps below 0.1\%. This is remarkable given the difficulty of the problem, which combines characteristics of the IRP and of the M-M PDP. These results are displayed in Table 3, where we present in columns “upper bound (without variable fixing)” and “upper bound (with variable fixing)” the upper bounds before and after applying the variable fixing phase, as well as the “improvement due to variable fixing (%).” Again, note that the latter solutions are valid for the general problem. The size of the vertex set $\mathcal{V}′$ is also listed: all instances but two contain 200 RATMs. These two exceptions resulted from the indivisibility of the number of RATMs (i.e., 6,377) over the 32 instances. In column “lower bound (with variable fixing)” we present the lower bound obtained when solving the problem after applying the variable fixing procedure, i.e., the constrained problem. We observe that these lower bounds are not valid for the general problem, which is less constrained. The column “optimality gap (%)” refers to the optimality gap between the upper and lower bounds computed after variable fixing. By increasing the number of periods $m$ in which the algorithm can decide when to serve the RATMs, we obtain a problem that turns out to be very similar to the original one, and the same solutions are obtained when $m = 1$ and $m = 2$.

We have also tested different cases by allowing the algorithm to visit RATMs that do not necessarily need a visit, but that could help reallocate cash throughout the system, thus avoiding the need to return cash to the depot, which incurs a cost. Whenever we set the number $f$ of close-by RATMs allowed to be visited to 1 or 2, but did not allow them to be visited in different periods, i.e., $m = 0$, the same solutions from the general

| Instance | $|\mathcal{V}′|$ | Upper bound (without variable fixing) | Upper bound (with variable fixing) | Improvement due to variable fixing (%) | Lower bound (with variable fixing) | Optimality gap (%) |
|----------|----------------|-------------------------------------|-----------------------------------|--------------------------------------|-------------------------------|-----------------|
| Inst-01  | 200            | 37,289                              | 25,698                            | 31.08                                | 25,642                        | 0.22            |
| Inst-02  | 191            | 34,076                              | 23,890                            | 29.89                                | 23,849                        | 0.17            |
| Inst-03  | 200            | 35,898                              | 25,594                            | 28.07                                | 25,576                        | 0.06            |
| Inst-04  | 200            | 33,962                              | 24,988                            | 26.42                                | 24,949                        | 0.15            |
| Inst-05  | 200            | 34,214                              | 25,420                            | 25.70                                | 25,394                        | 0.10            |
| Inst-06  | 200            | 31,837                              | 25,067                            | 21.26                                | 25,055                        | 0.05            |
| Inst-07  | 200            | 43,820                              | 25,190                            | 24.51                                | 25,181                        | 0.03            |
| Inst-08  | 200            | 33,028                              | 24,313                            | 20.54                                | 24,301                        | 0.02            |
| Inst-09  | 200            | 32,564                              | 23,663                            | 21.26                                | 23,644                        | 0.08            |
| Inst-10  | 200            | 29,865                              | 24,165                            | 23.29                                | 24,146                        | 0.00            |
| Inst-11  | 200            | 30,505                              | 24,381                            | 24.16                                | 24,362                        | 0.06            |
| Inst-12  | 200            | 30,684                              | 24,708                            | 25.30                                | 24,685                        | 0.09            |
| Inst-13  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-14  | 200            | 32,564                              | 24,831                            | 25.30                                | 24,816                        | 0.06            |
| Inst-15  | 200            | 28,123                              | 23,633                            | 27.33                                | 23,614                        | 0.08            |
| Inst-16  | 200            | 30,684                              | 24,381                            | 26.13                                | 24,362                        | 0.00            |
| Inst-17  | 200            | 31,620                              | 24,708                            | 23.29                                | 24,685                        | 0.00            |
| Inst-18  | 200            | 32,564                              | 24,831                            | 25.30                                | 24,816                        | 0.06            |
| Inst-19  | 200            | 30,505                              | 24,381                            | 24.16                                | 24,362                        | 0.09            |
| Inst-20  | 200            | 31,620                              | 24,708                            | 25.30                                | 24,685                        | 0.12            |
| Inst-21  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-22  | 200            | 32,564                              | 24,831                            | 25.30                                | 24,816                        | 0.06            |
| Inst-23  | 200            | 29,385                              | 24,708                            | 25.30                                | 24,685                        | 0.19            |
| Inst-24  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-25  | 200            | 31,620                              | 24,708                            | 25.30                                | 24,685                        | 0.00            |
| Inst-26  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-27  | 200            | 31,620                              | 24,708                            | 25.30                                | 24,685                        | 0.00            |
| Inst-28  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-29  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-30  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-31  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Inst-32  | 200            | 31,620                              | 25,070                            | 24.62                                | 25,050                        | 0.00            |
| Average  | 199            | 33,094                              | 25,114                            | 24.85                                | 24,859                        | 0.09            |
Although the results of this study have not been implemented at the time of this writing, we can provide some estimates of its benefits. According to the data to which we had access, the expected business cost savings is estimated at about €10.1 million per year only in the Netherlands, when compared with current practice, mostly due to the reduction in the number of required armored trucks. Since RATMs are expected to become a market standard in the near future, our work anticipates this development by providing cash supply chain parties the means to immediately improve the replenishment operations. In addition to significant cost savings, we emphasize that our methodology yields extra potential benefits, which are harder to quantify in monetary terms. Indeed, we observe that the maximum amount of money that can be carried by an armored truck is limited by its (very expensive) insurance policy. The larger the average amount of cash yields better upper bounds for all instances, with average improvements of 29.94% and attaining 47.17% in one case.

### 4.4. Estimated Benefits

Although the results of this study have not been implemented at the time of this writing, we can provide some estimates of its benefits. According to the data to which we had access, the expected business cost savings is estimated at about €10.1 million per year only in the Netherlands, when compared with current practice, mostly due to the reduction in the number of required armored trucks. Since RATMs are expected to become a market standard in the near future, our work anticipates this development by providing cash supply chain parties the means to immediately improve the replenishment operations. In addition to significant cost savings, we emphasize that our methodology yields extra potential benefits, which are harder to quantify in monetary terms. Indeed, we observe that the maximum amount of money that can be carried by an armored truck is limited by its (very expensive) insurance policy. The larger the average amount of cash

### Table 4 Summary of the Results When Changing the Parameter \( b \), Controlling the Inventory Levels of Extra RATMs

<table>
<thead>
<tr>
<th>Variable fixing parameters</th>
<th>Upper bound (without variable fixing)</th>
<th>Upper bound (with variable fixing)</th>
<th>Improvement due to variable fixing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f = 1, m = 1 )</td>
<td>b = 30</td>
<td>33,094</td>
<td>31,937</td>
</tr>
<tr>
<td></td>
<td>b = 50</td>
<td>33,094</td>
<td>30,802</td>
</tr>
<tr>
<td></td>
<td>b = 70</td>
<td>33,094</td>
<td>27,829</td>
</tr>
<tr>
<td>( f = 1, m = 2 )</td>
<td>b = 30</td>
<td>33,094</td>
<td>31,949</td>
</tr>
<tr>
<td></td>
<td>b = 50</td>
<td>33,094</td>
<td>31,944</td>
</tr>
<tr>
<td></td>
<td>b = 70</td>
<td>33,094</td>
<td>30,840</td>
</tr>
<tr>
<td>( f = 2, m = 1 )</td>
<td>b = 30</td>
<td>33,094</td>
<td>31,958</td>
</tr>
<tr>
<td></td>
<td>b = 50</td>
<td>33,094</td>
<td>31,934</td>
</tr>
<tr>
<td></td>
<td>b = 70</td>
<td>33,094</td>
<td>31,562</td>
</tr>
<tr>
<td>( f = 2, m = 2 )</td>
<td>b = 30</td>
<td>33,094</td>
<td>31,952</td>
</tr>
<tr>
<td></td>
<td>b = 50</td>
<td>33,094</td>
<td>31,941</td>
</tr>
<tr>
<td></td>
<td>b = 70</td>
<td>33,094</td>
<td>31,266</td>
</tr>
</tbody>
</table>

### Table 5 Best Results of the Branch-and-Cut Algorithm with the Clustering Procedure

<table>
<thead>
<tr>
<th>Instance</th>
<th></th>
<th>Upper bound (without variable fixing)</th>
<th>Upper bound (with variable fixing)</th>
<th>Improvement due to variable fixing (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inst-01</td>
<td>200</td>
<td>37,289</td>
<td>22,538</td>
<td>39.56</td>
</tr>
<tr>
<td>Inst-02</td>
<td>191</td>
<td>34,076</td>
<td>23,890</td>
<td>29.89</td>
</tr>
<tr>
<td>Inst-03</td>
<td>200</td>
<td>35,898</td>
<td>25,594</td>
<td>28.70</td>
</tr>
<tr>
<td>Inst-04</td>
<td>200</td>
<td>33,962</td>
<td>24,988</td>
<td>26.42</td>
</tr>
<tr>
<td>Inst-05</td>
<td>200</td>
<td>30,505</td>
<td>24,754</td>
<td>18.85</td>
</tr>
<tr>
<td>Inst-06</td>
<td>200</td>
<td>34,214</td>
<td>25,420</td>
<td>25.70</td>
</tr>
<tr>
<td>Inst-07</td>
<td>200</td>
<td>33,984</td>
<td>24,978</td>
<td>26.50</td>
</tr>
<tr>
<td>Inst-08</td>
<td>200</td>
<td>31,837</td>
<td>25,067</td>
<td>21.26</td>
</tr>
<tr>
<td>Inst-09</td>
<td>200</td>
<td>43,820</td>
<td>25,190</td>
<td>42.51</td>
</tr>
<tr>
<td>Inst-10</td>
<td>200</td>
<td>31,955</td>
<td>25,070</td>
<td>21.55</td>
</tr>
<tr>
<td>Inst-11</td>
<td>200</td>
<td>39,339</td>
<td>24,647</td>
<td>37.35</td>
</tr>
<tr>
<td>Inst-12</td>
<td>200</td>
<td>29,865</td>
<td>24,313</td>
<td>18.59</td>
</tr>
<tr>
<td>Inst-13</td>
<td>200</td>
<td>32,564</td>
<td>23,663</td>
<td>27.33</td>
</tr>
<tr>
<td>Inst-14</td>
<td>200</td>
<td>26,123</td>
<td>26,498</td>
<td>5.77</td>
</tr>
<tr>
<td>Inst-15</td>
<td>200</td>
<td>30,684</td>
<td>25,705</td>
<td>16.22</td>
</tr>
<tr>
<td>Inst-16</td>
<td>178</td>
<td>31,620</td>
<td>18,601</td>
<td>41.15</td>
</tr>
<tr>
<td>Inst-17</td>
<td>200</td>
<td>33,619</td>
<td>21,677</td>
<td>35.52</td>
</tr>
<tr>
<td>Inst-18</td>
<td>200</td>
<td>37,026</td>
<td>22,235</td>
<td>35.95</td>
</tr>
<tr>
<td>Inst-19</td>
<td>200</td>
<td>28,290</td>
<td>24,708</td>
<td>12.38</td>
</tr>
<tr>
<td>Inst-20</td>
<td>200</td>
<td>31,525</td>
<td>19,204</td>
<td>39.08</td>
</tr>
<tr>
<td>Inst-21</td>
<td>200</td>
<td>33,028</td>
<td>18,861</td>
<td>42.89</td>
</tr>
<tr>
<td>Inst-22</td>
<td>200</td>
<td>31,669</td>
<td>24,732</td>
<td>21.90</td>
</tr>
<tr>
<td>Inst-23</td>
<td>200</td>
<td>29,385</td>
<td>19,238</td>
<td>34.53</td>
</tr>
<tr>
<td>Inst-24</td>
<td>200</td>
<td>36,754</td>
<td>19,417</td>
<td>47.17</td>
</tr>
<tr>
<td>Inst-25</td>
<td>200</td>
<td>31,767</td>
<td>19,271</td>
<td>39.34</td>
</tr>
<tr>
<td>Inst-26</td>
<td>200</td>
<td>31,970</td>
<td>19,721</td>
<td>38.31</td>
</tr>
<tr>
<td>Inst-27</td>
<td>200</td>
<td>28,714</td>
<td>23,771</td>
<td>17.21</td>
</tr>
<tr>
<td>Inst-28</td>
<td>200</td>
<td>33,452</td>
<td>20,028</td>
<td>40.13</td>
</tr>
<tr>
<td>Inst-29</td>
<td>200</td>
<td>35,921</td>
<td>21,361</td>
<td>40.48</td>
</tr>
<tr>
<td>Inst-30</td>
<td>200</td>
<td>30,078</td>
<td>24,529</td>
<td>18.45</td>
</tr>
<tr>
<td>Inst-31</td>
<td>200</td>
<td>32,503</td>
<td>24,081</td>
<td>25.91</td>
</tr>
<tr>
<td>Inst-32</td>
<td>200</td>
<td>33,661</td>
<td>21,010</td>
<td>37.58</td>
</tr>
<tr>
<td>Average</td>
<td>199</td>
<td>33,094</td>
<td>23,196</td>
<td>29.94</td>
</tr>
</tbody>
</table>
being transported, the more expensive the insurance becomes. Our algorithm achieves a 54% reduction in the in-transit inventory, when compared with the traditional 1-M-1 structure. Not only will the insurance premium decrease substantially, but a less expensive truck (due to reduced armor) could be deployed, CO₂ emissions can be reduced, and losses due to thefts should significantly go down.

5. Conclusions
We have introduced, modeled, and solved an inventory-routing problem with pickups and deliveries. We have been successful in solving a difficult application of the problem arising in the optimization of distribution and inventory management of cash in recirculation ATMs in the Netherlands. The problem was first modeled as a very large-scale nonlinear mixed-integer program. After applying a clustering procedure and linearizing some of the constraints, including some valid inequalities and fixing the values of several variables, we have solved the problem by branch and cut. Different settings of the variable fixing procedure were tested, and we have shown which ones are able to provide a good trade-off in terms of simplification of the problem and upper bound values. After variable fixing, we were able to solve exactly or to near optimality 32 instances involving up to 200 vertices. This size is similar to that of the largest IRP or M-M PDP instances that have been solved in the past. Our problem is obviously more difficult, because it combines these two features. Since recirculation ATMs are expected to become a market standard in the near future, our paper anticipates this development by providing cash supply chain parties the means to immediately improve the replenishment operations and yield significant cost savings.

Acknowledgments
The authors thank Calcul Québec for providing computing facilities. The authors also thank the associate editor and the anonymous referees for providing valuable comments on this paper. This work was partly supported by the Canadian Natural Sciences and Engineering Research Council [Grants 39682-10 and 2014-05764]. This support is gratefully acknowledged.

References

van Anholt et al.: Inventory-Routing Problem with Pickups and Deliveries
Transportation Science 50(3), pp. 1077–1091, © 2016 INFORMS


