Joint Routing and Charging Problem of Electric Vehicles With Incentive-Aware Customers Considering Spatio-Temporal Charging Prices

Canqi Yao, Shibo Chen, Member, IEEE, Mauro Salazar, Member, IEEE, and Zaiyue Yang, Senior Member, IEEE

Abstract—This paper investigates the scheduling problem of a fleet of electric vehicles, providing mobility as a service to a set of time-specified customers, where the operator needs to solve the routing and charging problem jointly for each EV. We consider incentive-aware customers and propose that the operator offers monetary incentives to customers in exchange for time flexibility. In this way, the fleet operator can arrive at a routing and charging schedule with lower costs, whilst the customers receive monetary compensation for their flexibility. Specifically, we first propose a bi-level optimization model whereby the fleet operator optimizes the routing and charging schedule accounting for the spatio-temporal varying charging price, jointly with a monetary incentive to reimburse the delivery time flexibility experienced by the customers. Concurrently the customers choose their own time flexibility by minimizing their cost. Second, we cope with the computational burden coming from this nonlinear bi-level optimization model with an accurate reformulation approach consisting of the Karush-Kuhn-Tucker optimality conditions, a Big-M-based linearization method, and the zero duality gap of convex optimization problems. We thus convert the proposed problem into a single-level optimization problem, which can be solved by a strengthened generalized Benders decomposition method holding a faster convergence rate than the classic generalized Benders decomposition method. To evaluate the effectiveness of the proposed mathematical model, we carry out numerous simulation experiments by using the VRP-REP data of Belgium. The numerical results showcase that the proposed mathematical model can reduce the delivery fees for the customers together with the cost of operation incurred by the fleet operator.

I. INTRODUCTION

TO FACE a large increase in freight transportation demand while transiting to a carbon neutral society [1], fleets of electric vehicles (EVs) have been deployed to provide mobility as a service to a set of customers. The fleet operator (see Fig. 1) aims to make timely decisions on the routing schedules and the charging schedule of EVs, subject to specific pickup and delivery time windows. Meanwhile, the customers have some flexibility in terms of the pickup and delivery time, and can trade this flexibility for delivery cost reduction.

In practice, the strict schedules put limitations on the operational potential of the fleet, hence allowing for some time flexibility with respect to the desired schedule can significantly improve the overall operational performance of the entire system. To address this issue, the fleet operator can purchase such time flexibility from customers, and still obtain the larger overall revenues. From the perspective of the customers, if enough compensation is offered, it can be beneficial to provide time flexibility in exchange for a reduction of the delivery fee. In this context, to maximize the fleet operator’s revenues, the fleet operator aims to seek a good balance between the monetary incentives offered to the customers and the benefits coming from the time flexibility provided by the customers. Note that with the adoption of an incentive design strategy into the EVs’ routing problem (EVRP), there are two-fold benefits: (1) the customers will receive the delivery fee savings with the discount offered; (2) when the operator has an adjustable time window, more flexibility for the routing and charging schedule will be obtained resulting in the reduced operation cost. For this reason this paper investigates the EVRP with incentive-aware customers accounting for the decision of customers, and the monetary incentive mechanism design of the fleet operator.

Related work: This paper organizes the relevant literature in three parts, as follows: (i) monetary incentive mechanism design, (ii) the EVRP, and (iii) optimization methods in solving EVRP. In order to leverage the power of monetary incentives, a few researchers focus on designing incentive schemes to optimize the EVs’ charging problem and EVs’ routing problem [2], [8], [9], [10], [11], [12], [13]. In order to find advanced transportation management solutions,
Xiong et al. [12] developed an integrated and personalized traveler information scheme to stimulate an energy efficient mobility decision. Aiming to obtain a better routing and charging scheme, Diaz et al. [14] proposed an incentive scheme, which grants more space to the EV demand-aggregator to devise a better operation scheme by determining the optimal charging time. However, these incentive schemes investigated in previous papers are solely used in either routing or charging applications, and there is no research publication on the joint optimization of the EV routing decisions, charging schedules, and incentive mechanism design problem.

Due to the local zero-carbon emissions of EVs, the EVRP has recently drawn great attention from the academic community [2], [11], [15], [16], [17], [18]. In order to cope with the computational complexity stemming from the routing and charging problem of EVs, which is essentially a mixed-integer nonlinear programming problem, we proposed a computational-efficient method based on convex optimization [15]. In a similar manner, Boewing et al. devise a solution algorithm with polynomial time guarantees to obtain the optimal vehicle routing and charging schedules of a fleet of electric autonomous vehicles offering mobility-on-demand service [16]. However, these papers do not incorporate the underlying interaction between the customers and the fleet operator, which can greatly enhance the performance of the entire system by reducing the operation cost for the fleet operator and helping the customers save on delivery fees. Given the time-varying characteristics of electricity pricing, there are some works focusing on the electric vehicle routing and charging problem with a time-of-use electricity price [2]. To handle the EV routing problem with time windows considering time-variant electricity prices, Lin et al. [2] proposed the Lagrangian relaxation approach and the tabu search method to obtain near optimal solutions. However, an impractical assumption that EV cannot start the charging process at a recharging station before the start of the subsequent time period, is made in this paper. Besides, a high number of researchers focus on the EVPR with the node-based charging price [5], [15], [18], [19], [20], [21]. In contrast, there are few papers working on the EVRP with the spatio-temporal charging price, which is important in practice since the charging price differs with the charging location and charging time. For these reasons we aim to devise a mathematical model incorporating a spatio-temporal charging price without such an assumption that EV cannot start the charging process at a recharging station before the start of the following time period.

In this paper, we roughly classify the solution methodologies of solving EVRP as two categories, (a) heuristic optimization methods [2], [6], [22], and (b) exact optimization methods [3], [4], [5], [7], [15]. To solve EVRP at a reasonable computational cost, Schneider et al. [6] leverage a hybrid heuristics algorithm which consists of variable neighborhood search and Tabu search method. Since there is no guarantee either on the feasibility of solutions or on the optimality of solutions, Desaulniers [4] propose exact branch-price-and-cut algorithms to solve EVRP with time windows, and Calik et al. [3] present an efficient Benders decomposition method to solve the electric vehicle location-routing problem. According to the review paper [23], the heuristic optimization method can be used to solve the EVRP. However, these heuristic methods cannot guarantee the optimality of solutions. The convergence properties of heuristic methods are not always well understood, and determining the optimal parameter settings for a heuristic optimization method can be a difficult and time-consuming task [24], [25], [26]. Therefore, developing an accurate method with lower computational complexities would be of advantage in addressing the EVRP. By leveraging the Lagrangian relaxation method and generalized Benders decomposition method, we develop a decomposition algorithm which achieves a better performance on the computational cost, compared to the generalized Benders decomposition method and branch-price-and-cut algorithms. In Table I, we compare the related works in terms of the mathematical models and solution methodologies. In addition, the column headings stand for spatial charging price (SCP), temporal charging price (TCP), spatio-temporal charging price (STCP), customer model’s constraint (CC); LRA, VNS, BD, BPC, MBD, TS, and SGB represent Lagrangian relaxation approach, variable neighborhood search, Benders decomposition, branch price and cut, modified Benders decomposition, Tabu search, and strengthened generalized Benders decomposition.

In this paper, the mathematical formulation of an EVs’ fleet operator is extended to incorporate the spatio-temporal electricity price at each charging station. Due to the hierarchical relation between fleet operator and customers, bi-level optimization is employed which is equivalently reformulated as a single-level optimization problem. In practice, the customers (e.g., the e-commerce company, and the goods distributor) are prone to obtain the transportation service from the delivery companies (fleet operator) like FedEx and UPS, which can provide transportation as a service. Thus, considering the aforementioned business model and the privacy-preserving requirement of a fleet operator and customers, it is reasonable to decompose the resultant single-level optimization problem to subproblems which are individually solved by the freight.
transportation operator and the customers. To this end we proceed with the strengthened generalized Benders decomposition (SGBD) method.

**Contributions:**
- We present a novel bi-level optimization model. On the upper level, the fleet operator aims to solve the EVRP with incentive-aware customers and spatio-temporal electricity price, determines the monetary incentives offered to the customers in exchange for the customers’ time flexibility. On the lower level, considering the incentives, customers decide the optimal flexibility provided to the fleet operator in order to minimize their delivery costs.
- To further cope with the nonlinear terms in the operator problem of the bi-level model, we propose a set of exact transformation techniques consisting of the Karush-Kuhn-Tucker (KKT) optimality condition, the Big-M linearization method, and the zero duality gap of the convex optimization problem to transform the proposed bi-level model to a single-level mixed-integer programming (MIP) problem.
- To address the inherent complexities of the single-level MIP problem, and the privacy-preserving requirement of the customers and the fleet operator, we devise an SGBD which combines the merits of the generalized Benders decomposition (GBD) method and the Lagrangian dual decomposition method. We prove that the cuts from SGBD are much tighter than the cuts of GBD.
- A preliminary version of this paper was accepted by the 2022 American Control Conference [21]. In this revised and extended version, we propose a detailed mathematical model including spatio-temporal electricity prices. Furthermore, leveraging the hierarchical structure of the customer problem and the fleet operator problem, we devise a decomposition-based algorithm combining the benefits of the generalized Benders decomposition and the Lagrangian dual decomposition method which iteratively solves the resulting single-level problem. Finally, additional numerical results regarding algorithmic performance are presented and discussed.

**Organization:** The paper unfolds as follows: The system models and the mathematical model of the bi-level EVRP with incentive-aware customers and spatio-temporal electricity price are elaborated in Section II. In Section III, an exact reformulation technique to reformulate the bi-level optimization problem as a single-level MIP, is devised. To achieve a decomposed implementation, we devise a decomposition algorithm with stronger valid cuts in Section IV. Extensive numerical simulations are carried out in Section V to demonstrate the superiority of the proposed bi-level mathematical model. In Section VI, we draw the conclusions of this paper.

Note that the terms “requests” and “customers” are used interchangeably.

**II. MATHEMATICAL FORMULATION**

To characterize a practical scenario of the EVRP with incentive-aware customers, we devise a bi-level optimization model to capture the interaction between the customers and the fleet operator. In this section we present the fleet operator model and the customers model as optimization problems, minimizing the operational costs and the total inconvenience perceived by the customers, respectively. Finally, we combine these two optimization problems into a bi-level framework.

### A. Fleet Operator Model

1) The Routing Process Constraint: We define the relation between electric vehicles and customers with a directed graph as in our previous work [15]. We denote the transportation network as a directed graph $G(V, E)$, where $V = \{v_1, C, R, v_n\}$ comprises a start depot $v_1$, an end depot $v_n$, charging station nodes $C$, and nodes representing the customers set $R$. The set $E$ stands for the set of paths with $(i, j) \in E$ representing a path from node $i$ to $j$. We define $d_{ij}$ as the travel distance between node $i \in V$ and node $j \in V$. Similarly, $T_{ij}, e_{ij}$ represent travel time and energy consumption from node $i$ to $j$, respectively. The binary variable $x_{ij}^k$ shows whether vehicle $k$ is assigned to traverse path $(i, j)$. In order to capture the practical characteristics of the charging network and electric vehicles, the electric vehicle flow conservation constraints, battery energy constraints and visiting time constraints are clarified as follows:

Here, all EVs are subject to the vehicle flow conservation constraint: An EV entering a customer node has to leave the same customer node, as well as EVs starting at the start depot return back to the end depot after serving the requests. This is given by

$$\sum_{j \in V} x_{ij}^k - \sum_{j \in V} x_{ji}^k = b_i, \quad \forall i \in V, \quad k \in K, \quad (1)$$

where $b_{v_1} = 1$, $b_{v_n} = -1$, $b_i = 0$, $i \in C \cup R$.  

---

**TABLE I: COMPARISON BETWEEN RELATED WORKS**

| References   | Model  | Objective function | SCF | TCP | STCP | CC | Distributed optimization algorithm | Solution methodology |
|-------------|--------|--------------------|-----|-----|------|----|-----------------------------------|----------------------|
| Lin et al. [2] | MIP 4   | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | LRA, VNS             |
| Çalık et al. [3] | MIP 1, 2, 5 | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | BD                   |
| Desaulniers [4] | MIP 1, 3 | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | BPC                  |
| Yao et al. [5] | MIP 1, 2, 4, 5, 6 | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | MBD                  |
| Schneider et al. [6] | MIP 3   | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | VNS, TS              |
| Duman et al. [7] | MIP 3   | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | BPC                  |
| Our paper    | MINLP 1, 2, 4, 5, 6 | ✓                  | ✓   | ✓   | ✓    | ✓  | ✓                                 | SGBD                 |

a: 1: travel time; 2: charging time; 3: travel distance; 4: charging cost; 5: EVs’ usage cost; 6: revenue from serving customers; 7: discount on the delivery fee.
There is a constraint that each customer is served once, at most, by an EV as
\[
\sum_{k \in \mathcal{K}} \sum_{j \in \mathcal{V}} x_{ij}^k \leq 1, \quad \forall i \in \mathcal{R}.
\] (2)

The constraint is imposed, because a few transportation customers may not be served by EVs if the serving cost is higher than the benefit occurring during the delivery service. Besides, visiting time constraints are characterized by
\[
t_j \geq T_{ij} + g_i \cdot r_i^k + t_i - M(1 - x_{ij}^k),
\]
\[
\forall i \in \mathcal{C}, \quad j \in \mathcal{V} \setminus \mathcal{V}_1, \quad k \in \mathcal{K},
\] (3)
\[
t_j \geq T_{ij} + t_i - M(1 - x_{ij}^k),
\]
\[
\forall i \in \mathcal{R}, \quad j \in \mathcal{V} \setminus \mathcal{V}_1, \quad k \in \mathcal{K},
\] (4)
in which (3) specifies that the arrival time at the following customer node \(t_j\) should be longer than the total of the arrival time of previous customers \(t_i\), charging time \(g_i \cdot r_i^k\) in which \(g_i\) and \(r_i^k\) represent the time duration per unit kilowatt and the charging amount of EV \(k\) at charging station \(i\), respectively, as well as travel time \(T_{ij}\). Constraint (4) denotes that the arrival time of the subsequent request node \(t_j\) should be longer than the total of the arrival time of previous transportation requests \(t_i\), and travel time \(T_{ij}\). The constant \(M\) of (3) and (4) representing a constant with a large value (e.g., \(M = 10^6\)), is introduced to avoid the occurrence of bilinear terms. The same techniques also apply to (6) and (7). In addition, each customer is constrained by their own time windows, which can be expressed as follows:
\[
t_{ij}^l \leq t_j \leq t_{ij}^u + \delta_{ij}, \quad \forall j \in \mathcal{R}.
\] (5)

2) The Charging Process Constraint: In contrast to the frequently used spatially varying electricity price of the mathematical model in literature [15], [20], we extend the mathematical model to incorporate the time-varying electricity price which is partly addressed by Lin et al. [2] and Ferro at al. [18]. However, they both have a similar but impractical assumption that the EVs need to recharge their batteries at the start of the time slot. Under the circumstance of the time-varying electricity price, since the start time of the charging process is dependent on the charging and routing schedules, it is hard to characterize the charging cost, which is the product of the charging energy and the time-varying electricity price. Extending the mathematical model of Ferro et al. [18], an additional continuous variable \(r_i^k\) and two binary variables \(B_{ir}\), \(B_{ir}^s\) are used to precisely characterize the relation between discrete time slots \(\tau\) and the continuous arrival time of EVs \(t_i\), and to incorporate the spatio-temporal electricity cost.

The battery energy dynamics of the EVs at the charging station node is defined by
\[
-M(1 - x_{ij}^k) \leq -E_j^k + E_i^k + r_i^k - e_{ij}x_{ij}^k
\leq M(1 - x_{ij}^k), \quad \forall i \in \mathcal{C}, \quad j \in \mathcal{V} \setminus \mathcal{V}_1, \quad k \in \mathcal{K},
\] (6)

where \(r_i^k\) is the amount charged at charging node \(i\). Besides, the battery energy dynamics of the EVs at the customer node are characterized by
\[
-M(1 - x_{ij}^k) \leq -E_j^k + E_i^k - e_{ij}x_{ij}^k \leq M(1 - x_{ij}^k), \quad \forall i \in \mathcal{R}, \quad j \in \mathcal{V} \setminus \mathcal{V}_1, \quad k \in \mathcal{K},
\] (7)
in which \(e_{ij} = \phi d_{ij}\) denotes the amount of energy consumption from node \(i\) to node \(j\), and \(\phi\) denotes the battery energy consumption per unit distance. As for constraint (6) and (7), the Big-M linearization approach is used to reformulate the original nonlinear constraints as the linear constraints. We put lower and upper bounds on the battery energy level of EV \(k\), \(E_i^k\) as
\[
0 \leq E_i^k \leq E_{\text{max}}^k, \quad i \in \mathcal{V}, \quad k \in \mathcal{K},
\] (8)

where \(E_{\text{max}}^k\) is the battery capacity of each EV. The initial battery energy level of EVs is
\[
E_i^{k|_0} = E^{k|_0}, \quad \forall k \in \mathcal{K}.
\] (9)

In addition, the time horizon is divided as a set of time slots \(\Lambda = \{1, \ldots, \xi\}\) where \(\xi\) represents the maximum time slots, with the constant length \(\Delta \tau\). Two binary decision variables \(B_{ir}\), \(B_{ir}^s\) are introduced, defining whether charging station \(i\) recharges in time slot \(\tau\) or not, and whether charging station \(i\) starts to recharge in time slot \(\tau\) or not, respectively. Besides, the relation between \(B_{ir}\) and \(B_{ir}^s\) is captured by
\[
B_{ir}^s \geq B_{ir} - B_{ir}^s(-1), \quad \forall i \in \mathcal{C}, \quad \tau \in \Lambda,
\] (10a)
\[
\sum_{\tau \in \Lambda} B_{ir}^s \leq 1, \quad \forall i \in \mathcal{C},
\] (10b)
which specifies that the maximum number of charging processes occurring during the entire time slot \(\Lambda\) is one. Note that in practice the charging station could provide the charging service several times over the entire horizon; in other words, during the whole horizon multiple charging processes could take place.

Remark 1: To accommodate the scenario in which multiple charging processes occur during the whole horizon, the charging nodes are duplicated with \(N\) dummy nodes, which are connected by the corresponding charging nodes by setting the distance to 0 where \(N\) denotes the maximum number of charging processes.

Considering that we can represent the charging time in two different forms, namely, (a) as the product of charging amount and charging time per kilowatt hour \(r_i^k g_i\) and (b) the sum of all charging time slots over the entire time horizon \(\sum_{\tau \in \Lambda} B_{ir} \Delta \tau\), the relation between \(r_i^k g_i\) and \(\sum_{\tau \in \Lambda} B_{ir} \Delta \tau\) is given by
\[
r_i^k g_i \leq \sum_{\tau \in \Lambda} B_{ir} \Delta \tau, \quad \forall i \in \mathcal{C}.
\] (11)

The relation between discrete time slot \(\tau\) and continuous arrival time of EVs \(t_i\) is characterized by
\[
\sum_{\tau \in \Lambda} B_{ir} \tau \Delta \tau \leq t_i, \quad \forall i \in \mathcal{C}
\] (12a)
\[
t_i \leq \sum_{\tau \in \Lambda} B_{ir}^s (\tau + 1) \Delta \tau, \quad \forall i \in \mathcal{C}.
\] (12b)
3) The Objective Function: With the time flexibility $\delta_j$ allowed by the customers, the fleet operator optimizes the operational cost comprising of:
- the charging cost $\sum_{i \in C} \sum_{t \in T} \frac{p_{it} B_{it} \Delta t}{g_i}$,
- the EVs usage cost and delivery revenue $\sum_{k \in K} \sum_{i \in V} \sum_{j \in V} c_i x_{ij}^k$,
- the travel time $\sum_{k \in K} \sum_{i \in V} \sum_{j \in V} \omega_T T_{ij} x_{ij}^k$,
- the charging time $\sum_{k \in K} \sum_{i \in V} \sum_{j \in V} \omega_T^k g_i x_{ij}^k$,
- the total discount $\sum_{j \in R} q_j \delta_j^* \sum_{i \in V} \sum_{k \in K} x_{ij}^k$,

which stands for the cost paid for the customers that provide a flexible delivery time window. Here $\omega_T$ denotes the value of time. Besides, $c_i$ represents a unified cost vector capturing both the vehicle usage fee $c_v$ and the negative delivery revenue $D_i$.\footnote{Since the fleet operator would like to minimize the operational cost, we set the delivery revenue $D_i$ of the customers as a negative value.}

$$c_i = \begin{cases} D_i, & \text{if } i \in R \\ c_v, & \text{if } i = v_1. \end{cases} \tag{13}$$

To summarize, the fleet operator problem with incentive mechanism design is stated as follows:

**Problem 1 (Operator model):**

$$\min \sum_{i \in C} \sum_{t \in T} \frac{p_{it} B_{it} \Delta t}{g_i} + \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} (c_i + \omega_T T_{ij} + \omega_T^k g_i) x_{ij}^k + \sum_{j \in R} q_j \delta_j^* \sum_{i \in V} \sum_{k \in K} x_{ij}^k$$

s.t. (1) – (13).

**B. Customers Model**

We assume that customers aim to maximize the incentive received, and minimize the inconvenience coming from their provided time flexibility $\delta_j$. Besides, we quantify the inconvenience of the customers with a monotonically increasing convex function $I(\delta_j)$ [21]. Specifically, given the discount price $q_j$ of customer $j$ determined by the fleet operator, we formalize the customer problem as follows:

**Problem 2 (Customer Model):**

$$\min_{\delta_j \in \mathbb{R}} I(\delta_j) - q_j \delta_j,$$

s.t. $0 \leq \delta_j \leq \bar{\delta}_j \tag{14}$

Thereby, we define these non-negative dual variables $u_j$, $\sigma_j$ which are associated with the inequality constraints of Problem 2.

**C. Joint Scheduling Problem of the Fleet Operator and Customers**

With the fleet operator model and customers model, we formulate the routing and charging problem of EVs considering the monetary incentives as follows:

**Problem 3 (Joint Routing and Charging Model of the Fleet Operator and Customers):**

$$\min \sum_{i \in C} \sum_{t \in T} \frac{p_{it} B_{it} \Delta t}{g_i} + \sum_{k \in K} \sum_{i \in V} \sum_{j \in V} (c_i + \omega_T T_{ij} + \omega_T^k g_i) x_{ij}^k + \sum_{j \in R} q_j \delta_j^* \sum_{i \in V} \sum_{k \in K} x_{ij}^k$$

s.t. $\delta_j^* \in \arg \min_{\delta_j} \left\{I(\delta_j) - q_j \delta_j, 0 \leq \delta_j \leq \bar{\delta}_j \right\}, \forall j \in R.$

Compared with related studies of EVRP [2], [4], [5], [6], [15], [18], [20], [21], in order to characterize a practical problem comprising of the spatio-temporal price and incentive-aware customers, the proposed Problem 3 is much more complicated in terms of constraints, and the formulation of objective function. In particular, we incorporate additional constraints (10)-(12) and variables $(B_{it}, B_{it}')$ into the proposed model. Due to these extra nonlinear constraints, variables, and the bi-level structure, the proposed model (Problem 3) is hard to optimize, in comparison to those existing works solely considering either time variant charging price [2], [18] or spatial charging price [5], [15], [20]. As shown in Table I, the objective function of the proposed model consists of a wide range of operational costs such as the charging cost, the travel time, the charging time, the total discount, the EVs usage cost, and delivery revenue. Here the term related to delivery revenue is coupled with the customer optimization model (Problem 2), resulting in a joint bi-level optimization model (Problem 3). Regarding the inherent difficulties raised by this bi-level optimization problem, there are a few studies attempting to solve the bi-level optimization problem [27], [28], [29]. As an alternative to solving the bi-level optimization problem [30] with a proposed bisection-based iterative approach, we propose to equivalently reformulate the bi-level optimization problem as a single-level optimization problem.

### III. EQUIVALENT SINGLE-LEVEL REFORMULATION APPROACH

Due to the inherent NP-hardness brought by the bi-level optimization, we cannot obtain the global optimal solution within the polynomial computation time [31]. For this reason directly solving Problem 3 is impractical. To address this issue, we employ the KKT optimality conditions of the customers optimization problem to accurately reformulate Problem 3 as a single-level mixed-integer nonlinear programming (MINLP) problem. In addition, to simplify such a complicated MINLP, we propose an accurate linearization method, which uses the Big-M linearization method and the strong duality of the convex optimization problem.

**A. Reformulation of Customer Problem**

In order to handle the NP-hardness of the proposed bi-level model, by leveraging the strong duality of the customer
problem of Problem 3, we represent the customer problem as a convex problem via its KKT optimality conditions:

\[
\begin{align*}
\nabla \mathcal{I}(\delta_j) - q_j - \sigma_j + u_j & = 0, \quad \forall j \in \mathcal{R}, \quad (15a) \\
0 & \leq u_j \perp (\delta_j - \tilde{\delta}_j) \geq 0, \quad \forall j \in \mathcal{R}, \quad (15b) \\
0 & \leq \sigma_j \perp \delta_j \geq 0, \quad \forall j \in \mathcal{R}, \quad (15c)
\end{align*}
\]

where \( \perp \) is the complementarity operator. The stationarity conditions are specified in (15a), whilst the primal feasibility constraint, the dual feasibility constraint, and complementary condition are given by (15b) and (15c).

Note that there is still a nonlinear term \( q_j \delta_j \) in the objective function of Problem 3, and in the complementarity constraints (15b) and (15c). In order to cope with the computational burden brought by these nonlinear terms, we devise an exact linearization approach to represent such nonlinear terms in a computationally-efficient manner.

### B. Accurate Linearization Method of the Nonlinear Terms

In order to further reduce the computational complexity of Problem 3, an equivalent linearization method consisting of (i) introducing auxiliary binary variables to linearize the nonlinear complementary constraint, and (ii) exploiting the strong duality of the customer problem [32] to accurately linearize the nonlinear term \( q_j \delta_j \), is devised.

1) **Linearized Complementary Constraints**: To cope with the nonlinear and nonconvex complementary constraints (15b), and (15c), rendering the resulting problem difficult to solve, we devise an exact linearization approach which linearizes these hard constraints by introducing auxiliary binary variables \( \psi^1_j, \psi^2_j \) and a constant \( M \) with the sufficiently large value, resulting in the subsequent disjunctive constraints [33]:

\[
\begin{align*}
0 & \leq \tilde{\delta}_j - \delta_j \leq M \psi^1_j, \quad \forall j \in \mathcal{R}, \quad (16a) \\
0 & \leq u_j \leq M(1 - \psi^1_j), \quad \forall j \in \mathcal{R}, \quad (16b) \\
0 & \leq \delta_j \leq M(1 - \psi^2_j), \quad \forall j \in \mathcal{R}, \quad (16c) \\
0 & \leq \sigma_j \leq M(1 - \psi^2_j), \quad \forall j \in \mathcal{R}.
\end{align*}
\]

2) **Linearized Objective Function**: There is still a nonlinear term \( q_j \delta_j \), which is the product of two continuous variables, in the objective function of Problem 3. With the customer model in Problem 2, in which \( \mathcal{I}(\delta_j) \) is a convex function, the Lagrangian function can be obtained as shown below:

\[
\mathcal{L}(\delta_j, u_j, \sigma_j) = \mathcal{I}(\delta_j) - q_j \delta_j - \sigma_j \delta_j + u_j(\delta_j - \tilde{\delta}_j)
\]

\[
= \mathcal{I}(\delta_j) + (u_j - q_j - \sigma_j) \delta_j - u_j \tilde{\delta}_j.
\]

The dual function can be obtained as follows:

\[
g(u_j, \sigma_j) = \inf_{\delta_j} \mathcal{L}(\delta_j) + (u_j - q_j - \sigma_j) \delta_j = u_j \tilde{\delta}_j + \inf_{\delta_j} \phi(\delta_j) = -u_j \tilde{\delta}_j + \phi^*(\delta_j^*).
\]

For convenience, we denote \( \phi(\delta_j) = \mathcal{I}(\delta_j) + (u_j - q_j - \sigma_j) \delta_j \).

According to the strong duality of convex optimization problems, there is no duality gap between the objective value of lower-level Problem 2 and the objective value of its dual problem. Then the complicating nonlinear term \( q_j \delta_j \) can be exactly linearized as follows:

\[
\begin{align*}
\mathcal{I}(\delta_j) - q_j \delta_j &= -u_j \tilde{\delta}_j + \phi^*(\delta_j^*) \\
q_j \delta_j &= \mathcal{I}(\delta_j) + u_j \tilde{\delta}_j - \phi^*(\delta_j^*).
\end{align*}
\]

In addition, to handle the bilinear terms \( \omega x^T g_i x^j \) and \( \mathcal{I}(\delta_j) + u_j \tilde{\delta}_j - \phi^*(\delta_j^*) \) in the objective function of Problem 3, the continuous auxiliary variables \( n^1_{ijk}, n^2_j \) are introduced as

\[
\begin{align*}
n^1_{ijk} & \geq \omega x^T g_i x^j - M(1 - x^k_j), \quad \forall i \in \mathcal{V}, \; j \in \mathcal{V}, \; k \in \mathcal{K}, \quad (17a) \\
n^2_j & \geq \mathcal{I}(\delta_j) + u_j \tilde{\delta}_j - \phi^*(\delta_j^*) - M(1 - \sum_{k \in \mathcal{K}, i \in \mathcal{J}} x^k_j), \quad \forall j \in \mathcal{R}, \quad (17b)
\end{align*}
\]

3) **Equivalent Reformulation of the Bi-Level Optimization Model of Operator and Customers**: For the sake of convenience, we denote all variables by \( X = \{x^i_j, n^1_{ijk} \in \mathcal{K}, (q_j, \delta_j, u_j, \sigma_j, \psi^1_j, \psi^2_j, n^2_j) \in \mathcal{V}, (\omega x^k_j), P_i \in \mathcal{V}, (B^1_t, B^2_t) \in \mathcal{A} \} \). With the proposed equivalent linearization methods, we can accurately reformulate Problem 3 as a Problem 4:

**Problem 4 (Linearized Single-Level Optimization Problem):**

\[
\begin{align*}
\min_X & \sum_{i \in \mathcal{K}, i \in \mathcal{J}} \sum_{j \in \mathcal{V}} \sum_{\alpha \in \mathcal{A}} \Bigl( n^1_{ijk} + n^2_j + (\omega x^T g_i x^j) \Bigr) \\
& + \sum_{i \in \mathcal{C}, \tau \in \mathcal{A}} p_i (B^1_t \Delta t) \\
\text{s.t.} \; (1) & - (13), \quad (15a), \quad (16), \quad (17).
\end{align*}
\]

Commercial solvers (e.g., Gurobi, and CPLEX) can be used to solve Problem 4 directly. However, considering the privacy-preserving requirement of both the fleet operator and the customers, we propose a decomposition-based algorithm to decompose the single-level optimization problem into a master problem and subproblem, which are solved in an iterative fashion. The master problem is solved by the fleet operator, and the subproblem is solved by a cloud operator of all customers. We assume that a cloud operator can collect all information regarding the solution of the subproblem. To book a ride-hailing service online, in a real-life scenario, the customers need to upload their travel information to the online ride-hailing platform, like Uber. In return, the online ride-hailing service providers offer customers multiple travel options. Then the assumption aforementioned is reasonable.

## IV. OPTIMIZATION FRAMEWORK WITH DECOMPOSITION

Considering the independence of the EVs fleet operator and the customers, it is impractical to directly solve the resultant single-level Problem 4 in a centralized fashion. In the community of optimization, enormous efforts are put into designing efficient algorithms to solve mixed-integer programming problems [34], [35], [36], [37], [38], [39]. Due to the hierarchical relation between the fleet operator and the customers, we propose an optimization algorithm based on the SGBD, which combines the complementary merits
of the Lagrangian dual decomposition and the GBD. More specifically, by exploiting the merits of the Lagrangian dual decomposition method, we devise an MIP subproblem. With the newly formulated subproblem, valid Benders cuts are generated that are tighter than the classic Benders cuts derived from the GBD. Besides, the tightness is also rigorously proved in the subsection below.

A. Generalized Benders Decomposition Method

For the sake of simplicity, the linearized single-level Problem 4 is presented in compact form as follows:

$$\begin{align*}
\min_{X_{d}, X_{c}} & \quad c_{d}^{T} X_{d} + c_{c}^{T} X_{c} \\
\text{s.t.} & \quad A_{d} X_{d} + A_{c} X_{c} \leq b_{a} \quad (18a) \\
& \quad B_{d} X_{d} + G(X_{c}) \leq b_{b} \quad (18b) \\
& \quad D_{d} X_{d} \leq b_{c} \quad (18c) \\
& \quad X_{d} \in \mathbb{R}^{m}, \quad X_{c} \in \mathbb{R}^{n} \quad \text{(18d)}
\end{align*}$$

where $X_{c} = \{\{q_{jk}, \sigma_{jk}\}_{j \in V}, \{u_{j}, \tau_{j}\}_{j \in V}, \{r_{ik}, e_{ik}\}_{i \in U}\}$ and $X_{d} = \{\{x_{ij}, \sigma_{ij}\}_{j \in V}, \{\tau_{j}, \psi_{j}\}_{j \in V}, \{b_{i}, B_{i}^{T}\}_{i \in E}\}$ represent the set of continuous variables and the set of discrete variables, respectively. Besides, $A_{d}, A_{c}, B_{d}, D_{d}$ are the coefficient matrices, $b_{a}, b_{b}, b_{c}, c_{d}, c_{c}$ are the coefficient vectors, and $n, m$ represent the dimension of continuous variables and binary variables, respectively. In addition, (18a) denotes the linear constraint, (18b) denotes the nonlinear constraint, where $G(X_{c})$ represents a nonlinear function, and (18c) represents the integer related constraints.

1) The Mathematical Formulation of the Benders Master Problem: By adopting an auxiliary variable $\Theta$ serving as a lower bound of the objective value of the subproblem, the Benders master problem (BMP) containing integer variables $X_{d}$ and the related constraints (18c) can be formulated as follows:

Problem 5 (BMP):

$$\begin{align*}
\min_{X_{d}, \Theta} & \quad c_{d}^{T} X_{d} + \Theta \\
\text{s.t.} & \quad D_{d} X_{d} \leq b_{c} \\
& \quad X_{d} \in \mathbb{R}^{m}.
\end{align*}$$

2) The Mathematical Formulation of the Benders Subproblem: With the solution obtained from solving BMP $X_{d}^{*}$ and the local copies of BMP variables $Z_{d}$, the Benders subproblem (BSP) can be formalized thus:

Problem 6 (BSP):

$$\begin{align*}
\min_{X_{c}, Z_{d}} & \quad c_{c}^{T} X_{c} \\
\text{s.t.} & \quad A_{d} Z_{d} + A_{c} X_{c} \leq b_{a} \quad (19a) \\
& \quad B_{d} Z_{d} + G(X_{c}) \leq b_{b} \quad (19b) \\
& \quad D_{d} Z_{d} \leq b_{c} \quad (19c) \\
& \quad Z_{d} = X_{d}^{*} \quad (19d) \\
& \quad Z_{d} \in \mathbb{R}^{m}, \quad X_{c} \in \mathbb{R}^{n} \quad \text{(19e)}
\end{align*}$$

If the solution of Problem 6 ($\tilde{X}_{c}, \tilde{Z}_{d}$) is feasible, the following Benders optimality cut can be generated:

$$\Theta \geq c_{c}^{T} \tilde{X}_{c} + \zeta^{T} (X_{d} - \tilde{Z}_{d}), \quad (20)$$

where $\zeta$ is the dual variable of (19d).

As for the infeasible Problem 6, with two slack variables $S_{a}, S_{b}$, the feasibility subproblem (FSP) is formulated as follows:

Problem 7 (FSP):

$$\begin{align*}
\min_{X_{c}, Z_{d}, S_{a}, S_{b}} & \quad \mathbf{1}^{T} S_{a} + \mathbf{1}^{T} S_{b} \\
\text{s.t.} & \quad A_{d} Z_{d} + A_{c} X_{c} \leq b_{a} + S_{a} \quad (21a) \\
& \quad B_{d} Z_{d} + G(X_{c}) \leq b_{b} + S_{b} \quad (21b) \\
& \quad D_{d} Z_{d} \leq b_{c} \quad (21c) \\
& \quad Z_{d} = X_{d}^{*} \quad (21d) \\
& \quad Z_{d} \in \mathbb{R}^{m}, \quad X_{c} \in \mathbb{R}^{n} \quad \text{(21e)}
\end{align*}$$

With the solution of Problem 7 ($\tilde{S}_{a}, \tilde{S}_{b}, \tilde{Z}_{d}$), the Benders feasibility cut can be generated as follows:

$$0 \geq \mathbf{1}^{T} \tilde{S}_{a} + \mathbf{1}^{T} \tilde{S}_{b} + \Upsilon^{T} (X_{d} - \tilde{Z}_{d}), \quad (22)$$

where $\Upsilon$ is the dual variable of (21d).

B. Strengthened Generalized Benders Decomposition Algorithm

Considering that directly applying the GBD is a time-consuming task [40], we devise a novel decomposition method, namely, the SGBD, which iteratively strengthens the master problem with tight optimality and feasibility cut by leveraging the power of the Lagrangian dual decomposition method.

1) Strengthened Optimality Cut: There are a few papers developing the Benders optimality and feasibility cut generated at fractional nodes of the search tree during the early stage of the search process, which helps to obtain a better lower bound [36], [41], [42]. In order to derive the strengthened optimality cut, by introducing the dual variable $\zeta$ for (19d), the Lagrangian dual problem of (6) can be formulated as follows:

$$\begin{align*}
\max_{\zeta} & \quad \min_{X_{c}, Z_{d}} c_{c}^{T} X_{c} - \zeta^{T} (Z_{d} - X_{d}^{*}) \\
\text{s.t.} & \quad A_{d} Z_{d} + A_{c} X_{c} \leq b_{a} \quad (23a) \\
& \quad B_{d} Z_{d} + G(X_{c}) \leq b_{b} \quad (23b) \\
& \quad D_{d} Z_{d} \leq b_{c} \quad (23c) \\
& \quad Z_{d} \in \mathbb{R}^{m}, \quad X_{c} \in \mathbb{R}^{n} \quad (23d)
\end{align*}$$

The strengthened Benders optimality cut can be generated by enforcing $Z_{d} \in \mathbb{R}^{m}$ within the Lagrangian dual problem of (6), as illustrated in Lemma 1.

Lemma 1: Given the linear relaxation solution of BMP $X_{d}^{*} \in \mathbb{R}^{m}$, and the dual variable $\zeta \in \mathbb{R}^{m}$ from solving BSP, we can formulate the following mixed-integer programming subproblem

$$\begin{align*}
\min_{X_{c}, Z_{d}} & \quad c_{c}^{T} X_{c} - \zeta^{T} (Z_{d} - X_{d}^{*}) \\
\text{s.t.} & \quad A_{d} Z_{d} + A_{c} X_{c} \leq b_{a} \quad (24a) \\
& \quad B_{d} Z_{d} + G(X_{c}) \leq b_{b} \quad (24b) \\
& \quad D_{d} Z_{d} \leq b_{c} \quad (24c) \\
& \quad Z_{d} \in \mathbb{R}^{m}, \quad X_{c} \in \mathbb{R}^{n} \quad (24d)
\end{align*}$$

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
to derive the strengthened Benders optimality cut (25) with the solution $\tilde{Z}_d, \tilde{X}_c$ from solving (24).

$$\Theta \geq c^T \tilde{S}_c + \xi (X_d - \tilde{Z}_d). \quad (25)$$

Proof: The proof can be found in Appendix A. □

Compared with the optimality cut generated by the GBD, the tightness of the strengthened Benders optimality cut is quantified and rigorously proved in Theorem 1.

Theorem 1: Given the dual multipliers $\bar{\gamma}$ of (19d) from solving Problem (6), the strengthened Benders optimality cut (25) is parallel to the generalized Benders optimality cut (20) at least $\Xi \geq 0$ units tighter, where

$$\Xi = \min_{\chi \cup \tilde{Z}_d \in \mathbb{B}^m} (c^T \chi - \bar{\gamma}^T \tilde{Z}_d) - \min_{\chi \cup \tilde{Z}_d \in \mathbb{B}^m} (c^T \chi - \bar{\gamma}^T \tilde{Z}_d).$$

Proof: The proof is delivered in Appendix B. □

With the tightly strengthened Benders cut of the SGBD, a larger area will be cut compared to the cut of the GBD. Thus, fewer iterations are achieved by the SGBD.

2) Strengthened Feasibility Cut: As for the infeasible strengthened Benders subproblem, the strengthened feasibility subproblem (SFSP) of SGBD is formulated with three additional non-negative slack variables ($S_a, S_b, S_c$) as follows:

**Problem 8 (SFSP):**

$$\min_{X_c, Z_d, S_a, S_b, S_c} \bar{1}^T S_a + \bar{1}^T S_b + \bar{1}^T S_c - \bar{\gamma}^T (Z_d - X_d^t)$$

s.t. $A_d Z_d + A_c X_c \leq b_a + S_a$ \quad (26a)

$B_d Z_d + G(X_c) \leq b_b + S_b$ \quad (26b)

$D_d Z_d \leq b_c + S_c$ \quad (26c)

$Z_d \in \mathbb{B}^m$, $X_c \in \mathbb{R}^n$. \quad (26d)

With the solution of Problem 8 ($\tilde{S}_a, \tilde{S}_b, \tilde{S}_c, \tilde{Z}_d$), the strengthened feasibility cut can be generated as follows:

$$0 \geq \bar{1}^T \tilde{S}_a + \bar{1}^T \tilde{S}_b + \bar{1}^T \tilde{S}_c + \bar{\gamma}^T (X_d - \tilde{Z}_d). \quad (27)$$

With these strengthened Benders cuts obtained (25), (27), compared with the GBD, the performance of the proposed SGBD is theoretically improved, which will be evaluated below, in the numerical simulation results.

C. Implementation Details

In this subsection, the implementation details are clarified as follows:

- Fractional solution: To quickly derive valid cuts, we first solve the LP relaxation of the BMP (Problem 5) with classical cuts;

- Strengthened cut generation: The strengthened Benders cuts (25), (27) are generated in the following steps: (a) the value of dual multipliers can be obtained by solving the subproblem (24) or the feasibility subproblem (8); (b) with the dual multipliers, strengthened Benders cuts (25), (27) can be generated.

Besides, the master problem is solved by the fleet operator, and the strengthened subproblem is solved by a cloud operator of the customers. Thanks to the power of the strengthened optimality and feasibility cut, compared with the GBD, the proposed SGBD holds a faster convergence rate over the whole solution process.

V. SIMULATION EXPERIMENTS

In this section we demonstrate the effectiveness of the proposed mathematical model with an example of package delivery. Besides, we provide the details of the numerical data used.

A. Illustrative Parameter Settings

The VRP-REP data for Belgium [43], which consists of a 1000-node map with their geographic coordinates, out of which we choose a finite number of customers in a random fashion, is used in the following numerical experiments. As for the specifications of EVs, the battery capacity and recharging power are set as 90 kWh, 150 kW (fast charging), and 22 kW (slow charging) [44], respectively. In addition, the energy consumption per kilometer and the average speed of EVs are set as 0.24 kW·h/km and 60 km/h, respectively. Considering the 24-hour planning horizon, there are 288 time slots with 5-minute time intervals. According to the real time electricity price of PJM wholesale market [45], we obtain the electricity price of 288 time slots during 5 minutes for the numerical simulation. We set the usage cost of an EV $c_v$ as 99 $.

In order to capture a realistic scenario, the positive revenue gained from serving a request, is sampled from a truncated normal distribution with the mean and the standard deviation set as 9.05 $ and 5 $ [46], respectively. For the sake of convenience, we model the inconvenience function as a convex two-segment piecewise affine function with $\gamma_1, 2 = \{0, 1.5\}$ $/h$ and $\gamma_{1,2} = [0.01, -0.01]$ $ for every customer. Furthermore, we can use the different inconvenience function for each customer to incorporate the heterogeneous case (i.e., every customer has the different time sensitivity $\gamma_2$). The simulation experiments are run with 50 times, and we showcase the results with average values in the subsequent figures. We solve the optimization problems with python 3.8 on a workstation with an Intel Core i9-10980XE processor consisting of 36 CPUs of 3.00 GHz and 64 GB of memory. The used packages of Python consist of gurobipy, numppy, time, csv, scipy.io.

B. Performance Comparison

To evaluate the benefit of introducing monetary incentives both for the fleet operator and the customers, with different values of the maximum time flexibility $\tilde{\delta}_j$ (which we set to be the same value for all customers), and with different values of the time sensitivity $\gamma_2$, the optimal solution of Problem 4, and the solution of the incentive-free problem detailed in Appendix C, are compared and analyzed.

1) Benefits Brought by the Time-Flexibility $\tilde{\delta}$: In order to assess the impact of different lengths of time flexibility on the reduction of the operation cost, we conduct 3 sets of experiments with the different value of $\tilde{\delta}$, $\tilde{\delta} \in \{0.5, 1, 1.5\}$, and $\gamma_2 = 1.5$ $/h$. As illustrated in Fig. 2, the operation cost is decreased while the value of $\tilde{\delta}$ increases from 0.5 h to 1.5 h, demonstrating that the operation cost can be further reduced as the customers provide a larger time flexibility. In turn, a higher degree of time flexibility $\tilde{\delta}$ will enable customers to receive a larger monetary incentive, ultimately leading to a larger delivery fee saving, as showcased in Fig. 3. Thereby, the
maximum average operation cost reduction is over 5%, with the largest average delivery fee saving exceeding 5%.

2) Benefits Brought by the Different Time-Sensitivity $\gamma_2$: We conduct three sets of simulation experiments with $\gamma_2 \in \{1.5, 2.5, 5\}$ $$/h$$ and $\delta = 1.5$$h$$ to showcase the impact of the different degrees of time sensitivity of customers on the reduction of the operation cost. As shown in Fig. 4, the value of the average operation cost becomes larger with the increasing value of $\gamma_2$. Thus, as the customers pay more attention on the time flexibility, the fleet operator has to provide a higher monetary incentive in exchange for the time flexibility. Besides, we also conclude that even in the extreme case, it is always beneficial for the fleet operator to provide such incentives to acquire the flexibility of variable time windows. As illustrated in Fig. 5, a larger value of $\gamma_2$ results in a larger average fee saving obtained, exceeding 13%.

3) Comparison With the State-of-the-Art Optimization Methods: Since BPC has been used a lot to solve (E)VRPs [4], [7], [47], [48] demonstrating its effectiveness in solving (E)VRP, we compare the proposed SGBD with GBD and BPC in terms of the following two aspects, (i) the number of iterations, and (ii) the computation time. Thanks to the power of the strengthened Benders cuts, the number of iterations of the SGBD (SGBD-i) method will be lower than that of the GBD (GBD-i) method inherently, which has been illustrated in the right Y-axis of Fig. 6. Different from the solution framework of GBD, BPC shows that a set-partitioning model is linearly relaxed and the resulting relaxation problem is solved by column generation. Besides, additional valid inequalities are used to strengthen the LP lower bound, and integer solutions can be finally enforced by continuously branching [4]. Thus, we only compare the computation time between SGBD (GBD), and BPC. As shown in Fig. 6, BPC holds the longest computation time in comparison with GBD and SGBD, demonstrating the superiority of SGBD in computation performance. Consequently, compared with the GBD (GBD-t), the SGBD (SGBD-t), which is involved in the computation of the mixed-integer subproblem, still achieves better performances over all instances in terms of the computation time as illustrated in the left Y-axis of Fig. 6.

4) Comparison With the Mathematical Model of [2] in Charging Costs: The mathematical model of [2] cannot accommodate the spatio-temporal charging price [49], [50], thus we have to set the charging price of all charging stations as the same value by taking the average value of the spatio-temporal charging price over all charging stations. Then, after solving the resulting optimization model, we obtain the charging costs under time-of-use charging price and the optimal routing and charging schedules in which the actual charging cost with the spatio-temporal price can be procured. Hence, compared with the actual charging costs with the spatio-temporal price, the deviation of charging cost with the time-of-use price under 50 runs of simulation is shown in Fig. 7. The maximum deviation exceeds 200%, validating the ineffectiveness of the model of [2]. Contrary to the model
of [2], our model (Problem 1) is able to accommodate the spatio-temporal price.

5) The Scalability of SGBD: To evaluate the scalability of SGBD, the numerical experiments are conducted with the number of nodes and the number of paths increasing from 11, 22 to 151, 302. Table II shows the average run time, as well as the gap between upper and lower bound. The ε of Table II denotes the gap between upper and lower bound, and * from the second row indicates the computation time reaching the time limit of 2 hours.

VI. CONCLUSION AND DISCUSSION

In this paper, we presented a bi-level optimization model to formulate the EVRP with flexible time windows, incentive-aware customers, and a spatio-temporal varying electricity price. Besides, we proposed an accurate reformulation method, and a strengthened generalized Benders decomposition method combining the complementary advantages of the Benders decomposition method and the Lagrangian dual decomposition approach, to reformulate the original bi-level optimization model and solve the resulting single-level optimization problem, respectively. The simulation results demonstrated that allowing for the time flexibility and jointly optimizing the monetary incentives offered to the customers in exchange of the flexible time window can improve the overall operational cost of the fleet operator by up to 5%, at the same time customers could benefit from a reduction of over 13% of the fleet operator by up to 5%, at the same time the flexible time window can improve the overall operational efficiency. The simulation results demonstrated that the solving SGBD still involves the computation of MIP which requires massive computation resources and computation time, in the future we aim to solve the convex relaxation problem of the original optimization model (lower bound), and recover a feasible solution (upper bound) from the relaxed solution. By this method, a polynomial time solution method can be devised which can guarantee real-time implementation. In addition, we would like to investigate more complex customers’ models as well as uncertain travel time [54] and energy consumption [22].

APPENDIX A

PROOF OF LEMMA 1

Without loss of generality, we use \( \mathcal{X} := \{(X_c, Z_d)\mid (24a), (24b), (24c), X_c \in \mathbb{R}^m\} \). When the following condition holds for any \((X_d, \Theta)\),

\[
\Theta \geq \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} \left\{ c_c^T X_c, Z_d = X_d \right\},
\]

the strengthened Benders optimality cut (25) is valid. Given any \((X_d, \Theta)\) satisfying the above inequality, we have

\[
\Theta \geq \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} c_c^T X_c
\]

\[
\geq \max_{\xi \in \mathbb{R}^m} \left\{ \xi^T X_d + \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} \left( c_c^T X_c - \xi^T Z_d \right) \right\}
\]

\[
= \max_{\xi \in \mathbb{R}^m} \left\{ \xi^T (X_d - X_d^*) + \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} c_c^T X_c - \xi^T (Z_d - X_d^*) \right\}
\]

\[
= \max_{\xi \in \mathbb{R}^m} \left\{ \xi^T (X_d - Z_d) + c_c^T X_c \right\}
\]

\[
\geq \xi^T (X_d - Z_d) + c_c^T X_c,
\]

where the second line follows from weak duality of the Lagrangian dual problem and the fourth row follows from the optimality of \((\tilde{X}_d, \tilde{X}_c)\). Thus, the strengthened Benders optimality cut is valid.

APPENDIX B

PROOF OF THEOREM 1

Given the solution of Problem (5) \( X_d^* \) and \( \tilde{\xi} \in \mathbb{R}^m \),

\[
\Theta \geq \max_{\xi \in \mathbb{R}^m} \left\{ \xi^T X_d^* + \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} \left( c_c^T X_c - \xi^T Z_d \right) \right\}
\]

\[
\geq \tilde{\xi}^T X_d^* + \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} \left( c_c^T X_c - \tilde{\xi}^T Z_d \right)
\]

\[
\geq \tilde{\xi}^T X_d^* + \min_{\mathcal{X} \cup Z_d \in \mathbb{R}^m} \left( c_c^T X_c - \tilde{\xi}^T Z_d \right).
\]

Note that the second and third inequalities correspond to the strengthened and classic generalized Benders optimality cuts, respectively. These two optimality cuts are parallel as they have the same slope. Compared with the optimality cut obtained from solving the generalized Benders decomposition method, the tightness of the strengthened Benders optimality cut can be quantified by

\[
\Xi = \min_{\mathcal{X} \cup Z_d \in \mathbb{B}^m} \left( c_c^T X_c - \tilde{\xi}^T Z_d \right) - \min_{\mathcal{X} \cup Z_d \in \mathbb{R}^m} \left( c_c^T X_c - \tilde{\xi}^T Z_d \right).
\]

Due to the positive value of \( \Xi \), the derived optimality cut is \( \Xi \) tighter than the classic generalized optimality cut.

Fig. 7. The deviation of charging costs (time-of-use price) from the actual charging costs (spatio-temporal price).

TABLE II

THE SCALABILITY OF SGBD

| Instance | \(|\mathcal{X}| = 11 | \(|\mathcal{X}| = 51 | \(|\mathcal{X}| = 101 | \(|\mathcal{X}| = 151 | \(|\mathcal{X}| = 302 |
|----------------|----------------|----------------|----------------|----------------|----------------|
| Run time (s) | 2.06 | 121.47 | 837.35 | * | |
| \(\epsilon\) | 0 | 0.41% | 2.33% | 12.23% | |
APPENDIX C

ELECTRIC VEHICLE ROUTING PROBLEM WITHOUT INCENTIVES

\[
\begin{align*}
\min \quad & \sum_{i \in C} \sum_{\tau \in \Lambda} \frac{p_{i\tau} B_{i\tau} \Delta \tau}{g_i} + \sum_{k \in K} \sum_{i \in \mathcal{V}} (c_i) \\
& + \omega T_j x_i j + \alpha T_j k x_i j \\
\text{s.t.} \quad & \sum_{j \in \mathcal{V}} x_i j - \sum_{j \in \mathcal{V}} x_i j = b_i, \quad \forall i \in \mathcal{V}; k \in K \\
& b_{w1} = 1, b_{w2} = -1, b_1 = 0, \\
& \sum_{k \in K} \sum_{j \in \mathcal{V}} x_i j \leq 1, \quad \forall i \in \mathcal{R} \\
& t_j \geq T_j + g_i r_i^k + t_i - M(1 - x_i j), \quad \forall i \in \mathcal{C}, \\
& j \in \mathcal{V} \setminus v_1, k \in K \\
& t_j \geq T_j + t_i - M(1 - x_i j), \quad \forall i \in \mathcal{R}, \\
& j \in \mathcal{V} \setminus v_1, k \in K \\
& l_i^j \leq t_j \leq l_i^j, \forall j \in \mathcal{R} \\
& -M(1 - x_i j) \leq -E_j^i + E_i^j + r_i^k - e_i x_i j \leq M(1 - x_i j), \forall i \in \mathcal{C}, j \in \mathcal{V} \setminus v_1, k \in K \\
& -M(1 - x_i j) \leq -E_j^i + E_i^j - e_i x_i j \leq M(1 - x_i j), \forall i \in \mathcal{R}, j \in \mathcal{V} \setminus v_1, k \in K \\
& 0 \leq E_i^j \leq E_{k, \max}, \quad i \in \mathcal{V}, k \in K \\
& E_i^0 = E_{k, \max}, \quad \forall k \in K. \\
\end{align*}
\]

\[c_i = \begin{cases} 
D_i, & \text{if } i \in \mathcal{R} \\
C_i, & \text{if } i = v_1 
\end{cases}
\]

\[B_{i\tau}^k \geq B_{i\tau} - B_{i(\tau-1)}, \forall i \in \mathcal{C}, \tau \in \Lambda \]

\[\sum_{\tau \in \Lambda} B_{i\tau}^k \leq 1, \forall i \in \mathcal{V} \]

\[\sum_{\tau \in \Lambda} B_{i\tau}^k \tau \Delta \tau \leq t_i, \quad \forall i \in \mathcal{C} \]

\[t_i \geq \sum_{\tau \in \Lambda} B_{i\tau}^k (\tau + 1) \Delta \tau, \quad \forall i \in \mathcal{C} \]

\[r_i^k g_i \leq B_{i\tau}^k \Delta \tau. \]

ACKNOWLEDGMENT

The authors thank Dr. I. New, F. Paparella, and O. Borisboom for proofreading this paper.

REFERENCES

[1] M. Corporation. (2022). Roadmap to a Carbon Neutral Society. [Online]. Available: https://www.mitsubishicorp.com/jp/en-carbon-neutral/pdf/20211018.pdf

[2] B. Lin, B. Ghaddar, and J. Nathwani, “Electric vehicle routing with charging/discharging under time-variant electricity prices,” Transp. Res. C, Emerg. Technol., vol. 130, Sep. 2021, Art. no. 103285.

[3] H. Çalık, A. Oulamara, C. Prodhon, and S. Salhi, “The electric location-routing problem with heterogeneous fleet: Formulation and benders decomposition approach,” Comput. Oper. Res., vol. 131, Jan. 2021, Art. no. 105251.

[4] G. Desaulniers, F. Errico, S. Inrich, and M. Schneider, “Exact algorithms for electric vehicle-routing problems with time windows,” Oper. Res., vol. 64, no. 6, pp. 1388–1405, Dec. 2016.

[5] C. Yao, S. Chen, and Z. Yang, “Online distributed routing problem of electric vehicles,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 9, pp. 16330–16341, Sep. 2022.

[6] M. Schneider, A. Stenger, and D. Goecke, “The electric vehicle-routing problem with time windows and recharging stations,” Transp. Sci., vol. 48, no. 4, pp. 500–520, Nov. 2014.

[7] E. N. Duman, D. Taş, and B. Çatay, “Branch-and-price-and-cut methods for the electric vehicle routing problem with time windows,” Int. J. Prod. Res., vol. 60, no. 17, pp. 5332–5353, Sep. 2022.

[8] E. N. Narassimhan and C. Johnson, “The role of demand-side incentives and charging infrastructure on plug-in electric vehicle adoption: Analysis of U.S. States,” Environ. Res. Lett., vol. 13, no. 7, Jul. 2018, Art no. 074032.

[9] M. Salazar, D. Paccagnan, A. Agazzi, and W. P. M. H. Heemels, “Urgency-aware optimal routing in repeated games through artificial currencies,” Eur. J. Control, vol. 62, pp. 22–32, Nov. 2021.

[10] I. Kucukoglu, R. Dewil, and D. Cattrysse, “The electric vehicle routing problem and its variations: A literature review,” Comput. Ind. Eng., vol. 161, Nov. 2021, Art. no. 107650.

[11] R. Basso, B. Kulsar, and I. Sanchez-Diaz, “Electric vehicle routing problem with machine learning for energy prediction,” Transp. Res. B, Methodol., vol. 145, pp. 24–35, Mar. 2021.

[12] C. Xiong, M. Shahabi, J. Zhao, Y. Yin, X. Zhou, and L. Zhang, “An integrated and personalized traveler information and incentive scheme for energy efficient mobility systems,” Transp. Res. C, Emerg. Technol., vol. 113, pp. 57–73, Apr. 2020.

[13] B. Yildiz and M. Savelbershel, “Pricing for delivery time flexibility,” Transp. Res. B, Methodol., vol. 133, pp. 230–256, Mar. 2020.

[14] P. Diaz-Cachinero, J. I. Muñoz-Hernandez, and J. Contreras, “Integrated operational planning model, considering optimal delivery routing, incentives and electric vehicle aggregated demand management,” Appl. Energy, vol. 304, Dec. 2021, Art. no. 117698.

[15] C. Yao, S. Chen, and Z. Yang, “Joint routing and charging problem of multiple electric vehicles: A fast optimization algorithm,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 7, pp. 8184–8193, Jul. 2022.

[16] F. Boewing, M. Schiffer, M. Salazar, and M. Pavone, “A vehicle coordination and charge scheduling algorithm for electric autonomous mobility-on-demand systems,” in Proc. Amer. Control Conf. (ACC), Jul. 2020, pp. 248–255.

[17] J. Luke, M. Salazar, R. Rajagopal, and M. Pavone, “Joint optimization of autonomous electric vehicle fleet operations and charging station siting,” in Proc. IEEE Int. Conf. Intell. Transp. Syst. Conf. (ITSC), Sep. 2021, pp. 3340–3347.

[18] G. Ferro, M. Paolucci, and M. Robba, “Optimal charging and routing of electric vehicles with power constraints and time-of-use energy prices,” IEEE Trans. Veh. Technol., vol. 69, no. 12, pp. 14436–14447, Dec. 2020.

[19] S. Pourazarm and C. G. Cassandras, “Optimal routing of energy-aware vehicles in transportation networks with inhomogeneous charging nodes,” IEEE Trans. Intell. Transp. Syst., vol. 19, no. 8, pp. 2515–2527, Aug. 2018.

[20] T. Chen, B. Zhang, H. Pourbabak, A. Kavousi-Fard, and W. Su, “Optimal routing and charging of an electric vehicle fleet for high-efficiency dynamic transit systems,” IEEE Trans. Smart Grid, vol. 9, no. 4, pp. 3563–3572, Jul. 2018.

[21] C. Yao, S. Chen, M. Salazar, and Z. Yang, “Incentive-aware electric vehicle routing problem: A bi-level model and a joint solution algorithm,” in Proc. Amer. Control Conf. (ACC), Jun. 2022, pp. 4662–4670.

[22] S. Pelletier, O. Jabali, and G. Laporte, “The electric vehicle routing problem with energy consumption uncertainty,” Transp. Res. B, Methodol., vol. 126, pp. 225–255, Aug. 2019.

[23] B. Li, G. Wu, Y. He, M. Fan, and W. Pedrycz, “An overview and experimental study of learning-based optimization algorithms for the vehicle routing problem,” IEEE/CASS J. Autom. Sinica, vol. 9, no. 7, pp. 1115–1138, Jul. 2020.

[24] Z. Ning et al., “Online scheduling and route planning for shared buses in urban traffic networks,” IEEE Trans. Intell. Transp. Syst., vol. 23, no. 4, pp. 3340–3444, Apr. 2022.

[25] Y. Li, F. Chu, C. Feng, C. Chu, and M. Zhou, “Integrated production inventory routing planning for intelligent food logistics systems,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 3, pp. 867–878, Mar. 2019.

[26] L. Yue and H. Fan, “Dynamic scheduling and path planning of automated guided vehicles in automatic container terminal,” IEEE/CASS J. Autom. Sinica, vol. 9, no. 11, pp. 2005–2019, Nov. 2022.
