Bilevel Facility Location with Endogenous Queueing: An Application to EV Charging Network Design

Weiquan Wang[1], Amira Dems[2], Yossiri Adulyasak[1], Okan Arslan[1], and Jean-François Cordeau[1]

[1]HEC Montréal, Montreal, Canada, ✉ {weiquan.wang, yossiri.adulyasak,okan.arslan, jean-francois.cordeau}@hec.ca
[2]Hydro Québec Research Institute (IREQ), Varennes, Canada, ✉ dems.amira@hydroquebec.com

The rapid electrification of transportation requires strategic infrastructure planning that accounts not only for range anxiety but also for congestion at charging facilities. We consider a bilevel facility location problem where a decision maker (leader) invests in charging facilities to serve users (followers) traveling between origins and destinations along paths that require intermediate stops at facilities located by the leader. In this framework, the leader maximizes the total origin-destination flow served under a budget constraint, while the followers operate under range limitations and deviation tolerances. In this context, the users choose paths that maximize their utility (or minimize disutility), which is characterized by travel time, as well as waiting time (modeled as an M/M/c queuing system) and service time at the facilities along the path. We model the endogenous nature of these delays where EV path selections determine the charging demand at each node, which in turn affects queuing times that influence drivers' path choices. To better characterize users' rationale in path selection, we introduce the concept of $\epsilon-$optimality which allows users to select any feasible path within a bounded deviation from the optimum.

In collaboration with Hydro-Québec, we explore the modeling framework and solution algorithms for this problem in the context of EV fast-charging network design. The bilevel problem is modeled using arc-based and path-based formulations, which are reformulated into single-level mixed-integer linear programming (MILP) models through $\epsilon-$optimality strong duality and piecewise-linear approximations of queueing delays. Although these single-level reformulations can be solved by a commercial solver, they face scalability issues on larger networks. To tackle realistic instances, we propose several computational enhancements, including an enhanced length-bounded cut generation procedure and a logic-based Benders decomposition algorithm. The cut generation component is designed to efficiently identify relevant charging-feasible paths without enumerating all possibilities. The logic-based Benders decomposition separates location decisions from the congestion-aware drivers' path evaluation and iteratively solves the master problem using logic-based cuts generated from the subproblem.

Extensive experiments on large real-world networks from California and the case study in Québec demonstrate the scalability and effectiveness of the proposed approach. The decomposition algorithm achieves order-of-magnitude speedups over single-level MILP formulations and solves the majority of large-scale instances to optimality or near-optimality within two hours. Furthermore, benchmark comparisons on a related problem from the literature show that our proposed decomposition algorithm with the enhanced cut generation procedure consistently outperforms the algorithm in the literature on a similar bilevel problem. The numerical study also highlights that incorporating endogenous congestion is important for EV charging planning and that the proposed framework can support large-scale network investment analysis in practice.