DOI: 10.1111/mice.12498

ORIGINAL ARTICLE

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm

Bahman Madadi[1] | **Rob van Nes**[1] | **Maaike Snelder**[1,2] | **Bart van Arem**[1]

[1]Department of Transport and Planning, Delft University of Technology, Delft, The Netherlands

[2]TNO, Delft, The Netherlands

Correspondence
Rob van Nes, Department of Transport and Planning, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands.
Email: r.vannes@tudelft.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 438-15-161

Abstract

This paper presents a bi-level model to optimize automated-vehicle-friendly subnetworks in urban road networks and an efficient algorithm to solve the model, which is relevant for the transition period with vehicles of different automation levels. We formulate the problem as a network design problem, define solution requirements, present an effective solution method that meets those requirements, and compare its performance with two other solution algorithms. Numerical examples for network of Delft are presented to demonstrate the concept and solution algorithm performances. Results indicate that our proposed solution outperforms competing ones in all criteria considered. Furthermore, our findings show that the optimal configuration of these subnetworks depends on the level of demand; lower penetration rates of automated vehicles call for less dense subnetworks, and thereby less investments. Nonetheless, a large proportion of benefits are already achievable with low-density subnetworks. Denser subnetworks can deliver higher benefits with higher penetration rates.

1 | INTRODUCTION

Automated driving (AD) is a trend that will evolve over time, both in the market penetration rate of automated vehicles (AVs) and the level of automation. According to SAE International (2018), level 5 is the ultimate automation level with an unlimited operating design domain (ODD), yet levels 3–4 AVs have a limited ODD. This implies roads that can accommodate these vehicles must meet certain standards. On the other hand, level 5 AVs are not expected on the market in the near future (Shladover, 2018), and even after their market introduction, for at least several decades, a heterogeneous mix of traffic with vehicles of different automation levels and different ODDs will be inevitable.

By definition, having a limited ODD means there are certain types of situations based on the infrastructure that levels 3–4 AVs cannot safely handle (SAE International, 2018). However, the conditions for these situations (exact differences in ODDs) have neither been clearly defined by automobile manufacturers nor by researchers. Some studies have advocated the need for infrastructure adjustments for AD (Courbon et al., 2016; Nitsche, Mocanu, & Reinthaler, 2014; Zhang, 2013). An overview of more studies that suggest specific infrastructure requirements for AD is provided in Farah, Erkens, Alkim, and van Arem (2018).

On the other hand, analyses of autopilot disengagement reports occurring during AV tests in the United States show that road surface conditions (including poor markings) is one of the main causes of disengagements (Lv et al., 2018). According to Dixit, Chand, and Nair (2016), 9.98% of disengagements have happened due to road infrastructure (which is the second cause after system failure), 5% due to other road users, and 56.1% due to system failures which can also be associated with vehicle's disability to deal with the environment. Moreover, number of disengagements differ significantly based on road type (lower in motorways and higher in urban streets). According to Favarò, Nader, Eurich, Tripp, and Varadaraju (2017), 89% of the reported accidents involving an AV happened at an intersection, with 32% occurring in city roads. Finally, it is stated in Favarò, Eurich, and Nader (2018) that more than 10% of disengagements are attributed to external factors with infrastructure and other road users being the main categories. Less than 13% of all disengagements happened in motorways, freeways, and arterial roads combined, and the rest of the disengagements (around 87%) happened in interstate roads and urban streets.

This signals the need for attention to infrastructure and its influence on safe operation of AVs. Guiding the AD flows toward specific parts of the network that are inherently safer, and upgrading the infrastructure in these parts to meet the needs of AVs can assist in ensuring safety for all road users and further promoting AD as well as realizing network performance benefits in the transition period. This requires developing a vision for urban road networks and a strategic plan for its implementation. So far, dedicated lanes and links for AVs as well as dedicated AV zones have been suggested and briefly studied in the literature (Chen, He, Zhang, & Yin, 2016; Chen, He, Yin, & Du, 2017; Ye & Wang, 2018). We suggest an alternative network configuration for urban regions. We envisage that AD for levels 3–4 AVs will be possible only on certain designated roads which is consistent with SAE's definition of levels 3–4 automation. These roads must be selected based on safety and quality considerations. This means either they meet certain standards or they have the potential to meet these standards after reasonable adjustments. For the remaining roads, manual driving will be mandatory (although supported by various assisting driving automation systems such as collision avoidance systems). On these designated roads, AVs will operate on the same lanes as conventional vehicles (CVs). In order to facilitate safe and efficient AD in mixed traffic on these selected roads, investments are required to ensure that they meet the desirable design standards (e.g., machine-readable and uniform lane markings and road signs, high surface quality, digital maps, and I2V communication infrastructure); hence there will be trade-offs between these investments and the benefits they provide. This necessitates a network design approach to decide which roads should be selected to facilitate AD in the transition period, and to assess the impacts of this selection.

In this paper, we formulate this problem as a bi-level network design problem (NDP), define requirements for its solution methods, propose a new heuristic algorithm that meets those requirements, and compare its performance with two variations of a common algorithm in the literature for solving the NDP. Moreover, based on the findings of the model, we discuss practical considerations for implementation of this network configuration.

The rest of this manuscript is organized as follows: Section 2 provides a brief background for the problem; Section 3 introduces the concept of subnetworks for AD and the problem formulation; Section 4 describes solution algorithms; two case studies and numerical results are presented in Section 5; and Section 6 includes the discussion and concluding remarks.

2 | BACKGROUND: NETWORK DESIGN PROBLEM AND AUTOMATED DRIVING

Strategic investment decisions regarding road networks are often considered within the concept of NDP (Farahani, Miandoabchi, Szeto, & Rashidi, 2013; Yang & Bell, 1998) where the objective is to optimize certain performance indicators for a network (e.g., total travel cost [TTC]) considering the travelers' reactions to the network performance (e.g., route choices). This can be modeled as a bi-level leader-follower Stackelberg game in which the leaders (i.e., transport authorities) supply the transport infrastructure aiming at optimizing their objective function (e.g., TTC) and the followers (i.e., travelers) react with their travel choices to optimize their own objective function (e.g., individual travel costs). However, classical NDP formulations do not consider AVs and their impacts on transport system and travelers' choices. AVs are expected to have far-reaching impacts on various aspects of mobility (Fagnant & Kockelman, 2015; Milakis, van Arem, & van Wee, 2017; Shladover, Su, & Lu, 2012). Some of these impacts, namely, changes in road capacity, fuel efficiency, and value of travel time (VoTT) need to be considered in the NDP in order to accurately capture the travel impacts of AVs. In the following subsections, we discuss these changes and different approaches for modeling and incorporating them into the NDP.

2.1 | Capacity changes due to AD

AVs can improve transportation systems, especially when combined with vehicle connectivity. Cooperative Adaptive Cruise Control (CACC) is one of the major potential benefits of AVs (Shladover, 2016; Shladover, Nowakowski, Lu, & Ferlis, 2015). According to Nowakowski, O'Connell, Shladover, and Cody (2010), CACC with vehicle-to-vehicle (V2V) communication can increase lane capacity by reducing the driving time headway from 1.4 s to approximately 0.6 s. Various microsimulation studies have similar (although less severe) conclusions about the increase in lane capacity after the introduction of AVs; see, for instance, Mahmassani (2016), Shladover et al. (2012), and van Arem, van Driel, and Visser (2006). This has been utilized in some NDP studies that consider AVs via assumptions of increased capacity in presence of AVs (Chen et al., 2016; Chen et al., 2017; Ye & Wang, 2018) and some other studies of AV impacts using macroscopic traffic assignment models (Levin & Boyles, 2015; Madadi, van Nes, Snelder, & van Arem, 2019).

All abovementioned macroscopic studies have used static traffic assignment models. Another approach for incorporating capacity changes as well as several other changes in traffic flow characteristics (e.g., traffic flow stability and throughput) in presence of AVs with more accuracy is via dynamic traffic assignment (DTA) models and adjusted fundamental diagrams. In order to model the differences in backwards wave speed and lane capacity in shared roads for AVs and CVs, Levin and Boyles (2016) proposed a multiclass cell transmission model (CTM) that was utilized later in Patel, Levin, and Boyles (2016) to evaluate the impacts of improved capacity due to AD efficiency and reservation controls on arterial and motorway networks. Other researchers have developed and used DTA models to study possible changes in traffic flow dynamics caused by AVs as well. However, using DTAs for NDP is computationally very challenging.

2.2 | VoTT changes due to AD

For trips in automated mode, changes in VoTT are likely due to the possibility of performing other activities while driving. Although this is not fully established and there is no consensus in the literature yet, there are studies that claim AD will lead to a lower VoTT (Correia, de Looff, van Cranenburgh, Snelder, & van Arem, 2019; de Looff, Correia, van Cranenburgh, Snelder, & van Arem, 2018; Le Vine, Zolfaghari, & Polak, 2015; Milakis, Snelder, van Arem, van Wee, & Correia, 2017). This can be considered in the route choice component of macroscopic traffic assignment models via differences in generalized travel cost; see, for instance, Levin and Boyles (2015) and Madadi et al. (2019).

2.3 | Fuel efficiency changes due to AD

Higher fuel efficiency is another expected advantage of CACC (Rios-Torres & Malikopoulos, 2017; Shida & Nemoto, 2009; Shladover et al., 2015). This can affect travelers route choice behavior since generalized travel cost is the defining factor for route choice. This has been utilized in Madadi et al. (2019) and Puylaert, Snelder, van Nes, and van Arem (2018) to explore the travel impacts of AVs.

2.4 | Network configurations for AD

Despite the rapidly growing body of knowledge about various aspects of automated driving, optimal network configurations for AVs using NDP frameworks are still rare in the literature. Chen et al. (2016) modeled the optimal deployment of dedicated lanes for AVs over time as an NDP where the lower level included a multi-user class (MUC) deterministic user equilibrium (DUE) route choice with CVs and AVs as separate classes, and the upper level included optimal decisions regarding location, timing, and number of exclusive lanes for AVs. Another optimal network configuration for AVs using NDP was introduced in Chen et al. (2017) where the authors proposed exclusive AV zones for urban road networks; for the AV zones with no CV access, a system optimal route choice was suggested, and for the other parts of the network, an MUC equilibrium was applied. Last, Ye and Wang (2018) attempted to optimize urban road networks for AVs with a combination of exclusive AV links and congestion pricing for CVs using a bi-level NDP.

So far the only NDP studies that consider AD concepts, have suggested exclusive lanes, links, and zones for AVs. A more realistic network configuration for the transition period with a heterogeneous mix of vehicles was introduced in Madadi et al. (2019) where the authors suggested promoting AD in mixed traffic only in specific parts of the network where the infrastructure is upgraded to enable safe and efficient AD. This leads to a subnetwork within the urban road network which is suitable for AVs, yet it is also accessible for other vehicles. The concept was referred to as AD subnetwork. An MUC stochastic user equilibrium (SUE) route choice based on the path-size logit (PSL) model was developed to assess the impacts of this subnetwork; however, the choice of links to include in the AD subnetwork was a priori. A scenario-based approach was used to compare the performance of several network configurations in absence of an optimization concept or a bi-level model. In this study, we model the problem as a bi-level NDP and develop a quantitative solution method for optimal construction of these subnetworks.

3 | OPTIMAL SUBNETWORKS FOR AUTOMATED DRIVING

In this section, we outline AD subnetwork configuration as well as the problem formulation, and we suggest solution methods for the problem.

3.1 | The concept of AD subnetworks

The concept of AD subnetworks entails the vision that AD will be facilitated on a subset of existing road networks, called AD subnetwork, within which AD will be allowed and utilized everywhere. In this study, we further develop the concept by formulating it as a bi-level NDP and developing a solution algorithm for its optimal deployment.

First, we specify a set of feasible links in the network that are assumed to be safe for AD after some adjustments based on sustainable safety principles (Wegman & Aarts, 2006; Wegman, Aarts, & Bax, 2008). These links include all roads with flow (mobility) function and some roads with distribution function. They correspond to motorways, expressways, and main urban roads which have been shown to be the safest roads for AVs and include none or very few intersections, which are where most AV accidents and disengagements have happened (Favarò et al., 2017; Favarò et al., 2018).

Next, we assume certain adjustment costs based on road types to meet design requirements for AD. These requirements include having limited access, high quality (e.g. pavement, lane marking, traffic signs, and lights), segregated traffic (homogeneity of mass and speed for vehicles in each lane), and grade separated or clear at-grade intersections. Note that most roads considered feasible here, already meet many of the mentioned standards. For instance, motorways require minimum or no adjustments to meet the standards. Therefore, we have assumed different adjustment costs based on road types and their initial quality with roads that are higher in the road network hierarchy (i.e., motorways) requiring least amount of adjustments thereby having the least adjustment costs.

Finally, since investing on adjusting all the links in large networks can be infeasible and even unnecessary, the next step is to make an optimal selection (among feasible links) that guarantees the highest possible gain from the investment.

3.2 | Operational concepts and assumptions

It is assumed in this study that all vehicles are allowed everywhere in the network but AD is only possible for AVs within the AD subnetwork in mixed traffic conditions (i.e., on the same lanes with CVs). Outside the AD subnetwork, all vehicles drive manually.

3.3 | Optimal AD subnetwork construction as a bi-level NDP

We model the optimal AD subnetwork construction problem as a bi-level NDP where the upper level entails the choice of links to include in AD subnetwork in order to minimize total (generalized) travel cost plus total (discounted) adjustment cost (investment), and the lower level represents the user equilibrium route choice.

The following subsections introduce the problem formulation, but first, we provide some relevant definitions and explanations which are used throughout the manuscript. All the following definitions are demonstrated in Figure 1.

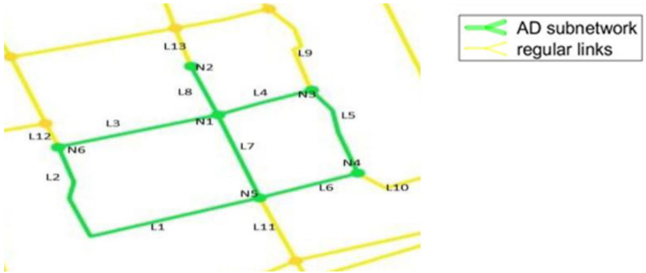


FIGURE 1 Definitions related to the AD subnetwork; L1–8 and N1–6 are included in the AD subnetwork which is connected and the rest of links and nodes represent a part of the underlying road network. N2–6 are boundary nodes but not N1; degree of N1 is 4 which is equal to the degree of its underlying node while the degree of N5 is 3 which is less than the degree of its underlying node (i.e., 4). L9–13 are boundary links; they are incident to a boundary node but not included in the AD subnetwork

Let $G(N, A)$ denote a directed graph representing the road network where N is the set of nodes (vertices) and A is the set of directed links (arcs) representing the road segments. $G_1(N_1, A_1)$ denotes the (directed) graph representing the AD subnetwork which is a subgraph of G and N_1 and A_1 represent nodes and links that are selected for the AD subnetwork. Conversely, $G_0(N_0, A_0)$ is the set of nodes and links in $G(N, A)$ that do not belong to $G_1(N_1, A_1)$. This represents the rest of the road network. $\hat{G}_1(N_1, \hat{A}_1)$ denotes an undirected graph that is obtained by replacing the directed links in G_1 with identical undirected links.

Definition 3.3.1. Any node incident to at least one link included in an AD subnetwork, is included in that AD subnetwork.

Definition 3.3.2. The degree of a node is the number of links incident to that node within the graph in which that node is included.

Definition 3.3.3. A boundary node n_1^* for an AD subnetwork represented by a directed graph G_1 is a node included in G_1 with a degree less than the degree of the corresponding node n in the underlying graph G (i.e., the graph representing the original road network).

Definition 3.3.4. A boundary link a^* for an AD subnetwork is a link incident to a boundary node n_1^* of the AD subnetwork but not included in the graph G_1 representing the AD subnetwork. Note that boundary nodes of an AD subnetwork are included in the subnetwork while the corresponding boundary links are not included in that subnetwork (hence the difference in the subscripts).

Definition 3.3.5. An AD subnetwork represented by $G_1(N_1, A_1)$ is called connected if replacing all of its directed links A_1 with undirected links $\hat{A}_1$ produces a connected (undirected) graph $\hat{G}_1(\hat{N}_1, \hat{A}_1)$. That is, for each two nodes in N_1, there is at least one path through links of $\hat{A}_1$ that connects these two nodes.

3.3.1 | Lower level problem: Multiclass equilibrium traffic assignment

For the lower level problem, we follow the formulation developed in Madadi et al. (2019) where the lower level is modeled as an MUC SUE problem based on the PSL. Since during the transition period, there will be a mix of different vehicle types in the network, we believe system optimal routing is unlikely. Two user classes (AVs and CVs) with different generalized travel cost functions are distinguished. On links within the AD subnetwork, AVs are assumed to have a lower passenger car unit (PCU or equivalently PCE) to account for the capacity changes due to AD, a lower VoTT because travel time can be used for other activities, and a lower driving cost per kilometer (which we call value of travel distance [VoTD]) due to fuel efficiency of AD. The generalized travel cost function is the summation of travel time multiplied by VoTT, and travel distance multiplied by VoTD. Link travel time is based on a modified Bureau of Public Roads (BPR) function where the total flow is a weighted sum of class-specific flows to capture the correlation between link capacity and the proportion of AVs on the link. Weighting is based on (adjusted) PCU values. Values used along with other key parameters used in this study are reported in Section 5.2.

Note that since static traffic assignments do not model traffic flow dynamics with a high level of details, capacity changes as a result of AD have been incorporated in static traffic assignment models in two different (simplified) manners; the first one is via assigning higher capacity to links that accommodate AVs which is more appropriate when considering dedicated lanes and links for AVs (Chen et al., 2016; Chen et al., 2017; Ye & Wang, 2018). The alternative is to consider an AV-flow-specific link travel time that is dependent on both cumulative flow of CVs and AVs on the link, and the ratio of CVs to AVs on the link (Levin & Boyles, 2015; Madadi et al., 2019). This approach is more appropriate when considering a mix of traffic (CVs and AVs) on links. Therefore, we have opted for the second approach here. A mathematical formulation is presented below. The notation used throughout this paper can be found in Table 1.

LLMP:

$$\begin{aligned}\min_{f_{m,a}} \quad Z_L = & \sum_m \frac{1}{\mu_m} \sum_{w\in W} \sum_{r\in R^w} F_m^{w,r} \ln F_m^{w,r} \\ & - \sum_m \frac{1}{\beta_m} \sum_{w\in W} \sum_{r\in R^w} F_m^{w,r} \ln PS_m^{w,r} \\ & + \sum_{m\in M} \sum_{a\in A_0} \int_0^{q_a} \theta_0 l_a + \eta_0 t_a(x)dx \\ & + \sum_{a\in A_1} \int_0^{q_a} \theta_0 l_a + \eta_0 t_a(x)dx \\ & + \sum_{a\in A_1} \int_0^{q_a} \theta_1 l_a + \eta_1 t_a(x)dx \end{aligned} \tag{1}$$

$$s.t. \quad q_a = \gamma_0(f_{0,a} + f_{1,a}), \quad \forall a \in A_0 \tag{2}$$

$$q_a = \gamma_0 f_{0,a} + \gamma_1 f_{1,a}, \quad \forall a \in A_1 \tag{3}$$

$$t_a(q_a) = t_a^0 \left[1 + \alpha_a \left(\frac{q_a}{\Lambda_a}\right)^{b_a}\right], \quad \forall a \in A \tag{4}$$

$$\sum_{r\in R^w} F_m^{w,r} = D_m^w, \quad \forall w \in W, \quad \forall m \in M \tag{5}$$

$$\sum_{w\in W} \sum_{r\in R^w} F_m^{w,r} \delta_{m,a}^{w,r} = f_{m,a}, \quad \forall a \in A, \quad \forall m \in M \tag{6}$$

$$F_m^{w,r} \geq 0, \quad \forall w \in W, \quad \forall m \in M, \quad \forall r \in R^w \tag{7}$$

$$PS_m^{w,r} = \sum_{a\in r} \left(\frac{l_a}{l_r}\right) \left(\frac{1}{\sum_{r\in R^w} \delta_{m,a}^{w,r}}\right), \quad \forall w \in W, \quad \forall m \in M, \quad \forall r \in R^w \tag{8}$$

where Equation (1) presents the objective function of the lower level problem (LLMP). The first term is related to stochastic route flows, the second related to the PSL, and the next three terms present different generalized travel costs for each class on each subnetwork. Equations (2)–(4) show how total flows and travel times are calculated. Equation (5) guarantees the demand is met for each class and each OD pair, Equation (6) projects route flows on corresponding links, Equation (7) prevents negative flows, and Equation (8) shows how PS penalties are calculated.

3.3.2 | Upper level problem: Optimal design of AD subnetworks

The upper level objective function to be minimized includes TTC and discounted total adjustment cost (TAC) for upgrading links. The decision variables are binary integers representing links to be included in the AD subnetwork. The lower level equilibrium conditions are treated as constraints for the upper level optimization problem. A mathematical representation of the optimal AD subnetwork design is introduced below.

ULMP:

$$\min_{I_a} \quad Z_u = TTC(I_a) + \frac{TAC(I_a)}{\sigma} \tag{9}$$

s.t.

$$\begin{aligned} TTC(I_a) = \sum_{a\in A} \Big\{ & (1 - I_a) \left[(\eta_0 \bar{t}_a + \theta_0 l_a) \left(\bar{f}_{0,a} + \bar{f}_{1,a}\right)\right] \\ & + I_a[(\eta_0 \bar{t}_a + \theta_0 l_a)\bar{f}_{0,a} + (\eta_1 \bar{t}_a + \theta_1 l_a)\bar{f}_{1,a}] \Big\} \end{aligned} \tag{10}$$

TABLE 1 Nomenclature

Notation	Explanation	Notation	Explanation
A	Set of all links a in the road network $A = A_0 \cup A_1$	W	Set of origin-destination (OD) pairs w
A_0	Set of links a not in the AD subnetwork	R^w	Set of routes between OD pair w
A_1	Set of links a that belong to the AD subnetwork	M	Set of user classes m (0: CVs, 1: AVs)
σ	Factor combining hourly to yearly travel cost conversion and discount rate over 10 years	$\delta_{m,a}^{w,r}$	Route-link incidence matrices (assignment maps)
l_a	Length of link a (kilometers)	D_m^w	Demand for OD pair w and class m
ac_a	Adjustment cost of link a	$PS_m^{w,r}$	Path-size penalty of route r between OD pair w and class m
η_m	VoTT for class m (0: CVs, 1: AVs)	β_m	Path-size correction parameter for class m
γ_m	PCU value of class m	μ_m	Logit choice model parameter for class m
θ_m	VoTD for class m	$F_m^{w,r}$	Flow of route r between OD pair w for class m
Z_u	Objective function of the upper level mathematical problem (ULMP)	Z_L	Objective function of the lower level mathematical problem (LLMP)
I_a	Binary decision variable (ULMP) taking the value of 1 for links selected for AD subnetwork (A_1), and zero for other links (A_0)	$\overline{f}_{m,a}$	Equilibrium flow of class m on link a (decision variable of LLMP)
$TTC(I_a)$	Total travel cost based on the value of I_a	$t_a(q_a)$	Travel time on link a based on total flow
$TAC(I_a)$	Total adjustment cost based on the value of I_a	q_a	Total flow of links a (PCU-equivalent)
n^p	Population size	α_a, b_a	BPR function parameters for link a
n^g	Number of generations	Λ_a	Capacity of link a
m_i	Merging interval (ELS)	$\lvert p_{\hat{G}_1}^{s,t} \rvert$	Number of paths between source node s and sink node t traversing $\hat{G}_1$
n_c^p	Number of candidates (ELS)	n_κ	Number of connected components in graph
c_f	Crossover fraction (GA)	w_κ	Penalty weight for each disconnected part in a graph
n_e^p	Elite size (GA)		

$$TAC(I_a) = \sum_{a \in A} I_a ac_a \tag{11}$$

$$I_a(1 - I_a) = 0, \quad \forall a \in A \tag{12}$$

$$\left| p_{\hat{G}_1}^{s,t} \right| \geq 1, \quad \forall s, t \in N_1 \tag{13}$$

where $\overline{t}_a$, $\overline{f}_{0,a}$, and $\overline{f}_{1,a}$ are implicitly defined by solving the lower level mathematical problem (LLMP). Equation (9) introduces the objective function of ULMP which is a summation of *TTC* and discounted *TAC*. Note that another representation of the objective function in Equation (9) would be $\lambda TTC(I_a) + \frac{TAC(I_a)}{(1+\pi)^{\tau-1}}$, where π is the discount rate and λ converts the hourly travel cost to a yearly rate; however, for convenience of calculations, both terms were divided by λ, and $\lambda(1 + \pi)^{\tau-1}$ was replaced by σ. Equation (10) shows how *TTC* is calculated according to the link flows obtained in equilibrium, and Equation (11) shows how *TAC* is calculated based on link adjustment costs and the value of decision variables. Equation (12) represents the binary decision variables that take the value of one if the link is included in the AD subnetwork and zero otherwise. Equation (13) ensures that the AD subnetwork represented by G_1 is connected (according to Definition 3.3.5). Solution algorithms are discussed in the following section.

4 | SOLUTION ALGORITHMS

4.1 | Requirements and key performance indicators

To solve the lower level mathematical problem (LLMP), the MUC extension of the method of successive averages (MSA) algorithm introduced in Wu, Florian, and He (2006) is used. Since the PCU-based weights used here for calculating total flow is similar to the approach used to account for trucks on links, LLMP has the same structure and mathematical properties as the multi-class mixed car and truck traffic network equilibrium problem. Therefore, it can be solved by MUC MSA algorithm mentioned earlier that is shown to be suitable for such problems. For more details, the reader is referred to Madadi et al. (2019) and Wu et al. (2006).

The upper level problem, on the other hand, is an NP-hard problem that is non-convex and is known to be very challenging to solve, especially, when the lower level problem is an MUC SUE and the upper level decision variables are discrete. Several solution methods have been suggested in the literature. The reader is referred to Magnanti and Wong (1984) and Yang and Bell (1998) for reviews; however, within the last two decades, stochastic search and approximation heuristics have become the most preferred methods among transport researchers to solve NDPs (Migdalas, 1995; Poorzahedy & Rouhani, 2007), especially NDPs with discrete decision variables (DNDP). A more recent review study by Farahani et al. (2013) shows that Genetic Algorithms (GA) introduced in Holland (1975) and elaborately reviewed in Goldberg (1989) is one of the most commonly used methods for dealing with the complexity of DNDPs in recent years. Chakroborty (2003) concludes that GA-based procedures are highly effective for urban transit NDP which is a DNDP variation. This is perhaps due to their flexibility in modeling the underlying problem and their capability to produce nearly optimal solutions in reasonable times for large-scale problems. However, specific characteristics of AD subnetworks impose additional requirements on the solution algorithm that common GAs do not meet, such as graph connectivity requirement.

In this study, three major criteria for a suitable solution algorithm are considered, namely, effectiveness (degree of optimality), efficiency (computation time), and design quality (generating connected subnetworks). The first two criteria are relevant for the solutions of all DNDP variants. On the contrary, generating connected subnetworks has never been a requirement for DNDPs in the past. Therefore, the existing DNDP solution methods do not necessarily meet this criterion. However, in order to envisage an effective subnetwork within a road network that serves the mobility of AVs, it is crucial to guarantee at least a certain degree of connectivity within this subnetwork. Otherwise, switching frequently between automated and manual driving modes can compromise, most of all, safety, but also utility and efficiency gains of AD.

Since origins and destinations (centroids) are not located on feasible links for the AD subnetwork, guaranteeing a strongly connected subnetwork that provides accessibility from all origins to all destinations within the subnetwork is impossible, which is in fact due to the limited ODD of levels 3–4 AVs. Yet, when a connected subnetwork for AD exists, AVs can start manually, reach the subnetwork and proceed in automated mode, and switch to the manual mode again when they approach their destination. In this manner, a safe and efficient navigation within the subnetwork in automated mode without disruptions is ensured for the middle part of the trip, which can be a large proportion of the trip. Then the design quality of an AD subnetwork refers to being connected, as defined in Definition 3.3.5, mathematically expressed in Equation (13), and illustrated in Figure 1. Moreover, we discuss the practical implications of this concept for optimal designs in Section 5.3.2.

To introduce connectivity to the solution algorithm, different approaches can be considered. One approach is considering a penalty for switching from automated to manual driving mode in the utility function used in route choice; however, that can easily make the problem intractable since it requires a tremendous number of graph connectivity tests (i.e., one test per route for each fitness function evaluation). An alternative approach is to consider a penalty for disconnected designs. This is feasible since the number of connectivity checks are considerably less than the first option (one for each fitness function evaluation). Still, this may not be an efficient approach. Another alternative is to consider connectivity as a requirement and avoid disconnected designs altogether. Yet there is no solution method in NDP literature that can do this. Therefore, in this study, we embarked on developing a new heuristic solution method (i.e., an evolutionary local search [ELS] algorithm) that produces connected designs efficiently. To demonstrate its performance, we also considered two alternative algorithms: a GA and a modified GA (MGA) that uses a penalty function to avoid disconnected designs based on the second approach mentioned earlier. In the following sections, we describe these three algorithms.

4.2 | Genetic algorithm

Genetic algorithms have become one of the favorite methods to solve DNDPs. They essentially make successive improvements over several generations of populations, each containing a set of strings (individual solutions) until finding a desirable solution. They use three general operations, namely, reproduction, crossover, and mutation.

Reproduction is the process of selecting individuals that contribute to the next generation (parents). This process is random but biased based on individuals' fitness usually via a roulette wheel selection with slot sizes proportional to the individual's fitness value. Crossover operation includes constructing a mating pool based on roulette wheel, randomly matching individuals and swopping a portion of their characteristics. Mutation includes random perturbations on individuals to preserve diversity. Elitism is sometimes preferred in using GAs which includes selecting a certain number of individuals with best fitness (elites) and passing them to the next generation. This ensures that the good features in early generations survive to the later generations. A detailed explanation of GA operations is provided in Goldberg (1989).

In this study, each individual design (chromosome) represented by I_a is coded as a binary string of bits, 0 and 1 (genes) with a length equal to the number of feasible links for the subnetwork where each link is represented by a 1 if it is included in the subnetwork (A_1) and a 0 otherwise (binary or bit string coding). Mutation is applied to 1% of genes in each chromosome and mutation genes are uniformly distributed over the range of genes. A multiple-point crossover function

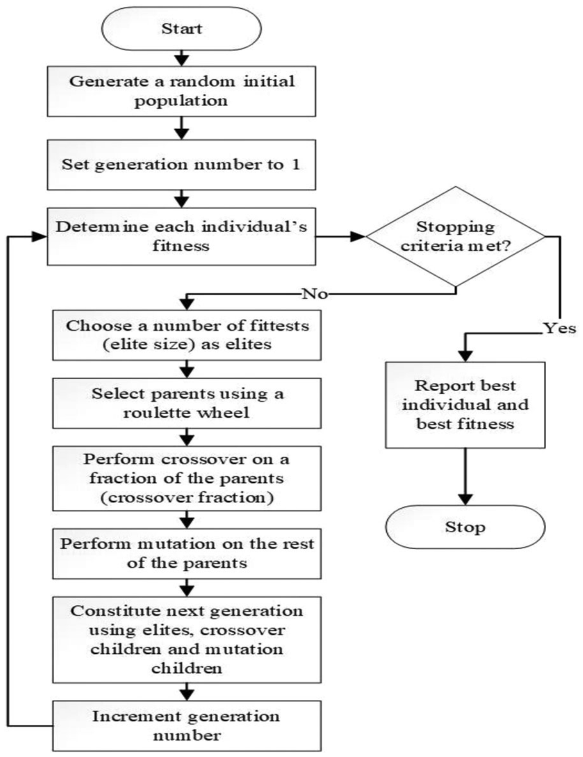


FIGURE 2 GA procedure for optimizing AD subnetwork

(sometimes referred to as uniform crossover) is used where each gene has an equal chance of coming from either parent. Fitness value in this study is the value of Z_u for each individual solution (Equation (9)) where the equilibrium flows therein are obtained from solving LLMP.

The GA procedure used in this study is demonstrated in Figure 2. The number of offspring produced by each operation is determined by parameter values that define elite size (number of elites) and crossover fraction (are reported in Section 5.2).

4.3 | Modified genetic algorithm

Since GAs generally are not tuned to produce connected designs, we use a modified GA (MGA) that penalizes disconnected designs in its fitness function evaluation in order to achieve connected optimal designs. The procedure of this MGA is similar to GA described in Section 4.2 and depicted in Figure 2; the only difference here is that a penalty term $w_\kappa\ (n_\kappa - 1)$ is added to the fitness function to aid the algorithm in finding connected designs. n_κ denotes the number of connected components and w_κ represents the penalty for each disconnected part. Note that in a connected graph the number of connected components is one. Therefore, the penalty term for connected graphs equals zero and with each disconnected part, the penalty increases. In this manner, the algorithm gets some feedback through improvements in fitness when moving from graphs with many separate connected components towards graphs with less connected components until (ideally) it reaches a graph with only one connected component. It is shown in Section 5 that this approach is effective yet not necessarily efficient. Then instead of Equation (9), MGA uses Equation (14) described below for its fitness function evaluations:

$$Z_u = TTC(I_a) + \frac{TAC(I_a)}{\sigma} + w_\kappa(n_\kappa - 1) \tag{14}$$

4.4 | Evolutionary local search algorithm

It is acknowledged that testing connectivity of graphs is a computationally burdensome task (Even & Tarjan, 1975; Farahani & Miandoabchi, 2012). Besides, AD subnetworks are constructed from scratch. That gives us the opportunity to start with an initial (simple) connected graph and extend the graph only using operations that preserve connectivity; hence, a weakly connected graph is guaranteed without undertaking the burden of testing for connectivity at each iteration. On the other hand, hybrid metaheuristics can outperform single heuristics when tailored to a specific problem using the information available about the problem (Poorzahedy & Rouhani, 2007). Our ELS algorithm is inspired by local search algorithms such as hill climbing (step 2) that are generally efficient and evolutionary algorithms such as GAs (step 3) that have been proven effective for DNDP. It combines ideas from mentioned metaheuristics and takes advantage of the problem structure (building subnetworks from scratch) to generate connected AD subnetworks in an efficient manner.

4.4.1 | ELS steps

ELS algorithm procedure consists of four major steps. Step 1 is initiation, step 2 is evolution to the next generation (i.e., adding links), step 3 is the merging process, and step 4 is the check for stopping criteria (i.e., termination). These steps are listed in detail below and depicted in Figure 3.

Step 1. Initiate n^p designs (subgraphs or individuals) for AD subnetwork in parallel, each one including one feasible link selected based on a roulette wheel where the odds are proportional to link capacities (this will favor links with higher capacities for initial designs).

Step 2. For each design:

Step 2.1. Find boundary nodes and boundary links

Step 2.2. Construct a set of n_c^p candidate new designs by adding a boundary link based on the roulette wheel selection (mentioned in step 1) to the existing design

Step 2.3. For each candidate new design:

Step 2.3.1. Measure fitness (evaluate the ULMP objective function Equation (9) by solving LLMP)

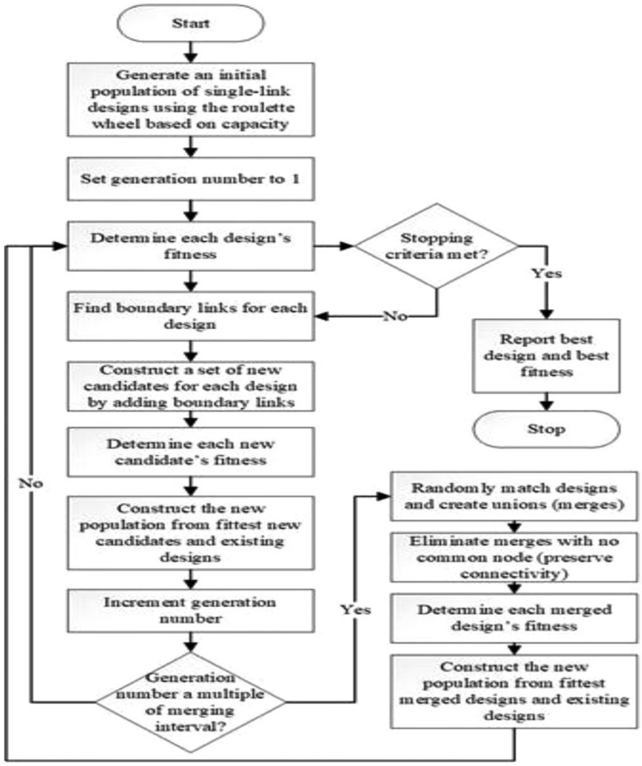


FIGURE 3 ELS procedure for optimizing AD subnetwork

Step 2.4. Replace the design with the relevant candidate new design that has the best fitness if it also exceeds the existing design in fitness.

Step 3. If the number of generations is a multiple of the merging interval m_i:

Step 3.1. Randomly match half the designs with the rest

Step 3.2. Eliminate the matches that do not have at least one node in common (to preserve connectivity)

Step 3.3. Construct new candidate designs by creating unions of matched designs (referred to as merged designs)

Step 3.4. Measure the fitness of each merged design

Step 3.5. Construct the new set of n^p designs by selecting the fittest n^p designs among the existing designs and the merged designs.

Step 4. Stop and return the fittest design if the stopping criterion is met (i.e., no improvement for five consecutive generations or no new candidate link), else go to step 2.

Note that while both ELS and GA are population-based algorithms, they have two fundamental differences. GAs start with a random number of links selected for the subnetwork and iteratively add and remove links to reach better designs. ELS starts from designs with one link and grows the subnetwork only with the addition of links that improve the fitness till reaching the optimal design. Second, GAs proceed with accepting disconnected designs as feasible solutions while reducing the probability of having disconnected designs via penalties (our MGA uses such a penalty). On the other hand, all ELS operations are designed to preserve connectivity: ELS starts with one link (which is a connected graph) and moves from one connected design to another via operating on boundary links and a conditional merging procedure that preserves connectivity.

5 | CASE STUDIES AND NUMERICAL RESULTS

5.1 | Case study 1: Synthetic network

In this section, a small synthetic network with 9 nodes and 24 links (12 pairs) is used as a proof of concept. We have considered local roads infeasible for AD subnetwork. Demand is assumed to be 280 v/hr from each node to any other node. The adjustment cost is discounted over 10 years ($\tau = 10$) (assumed effective life time for the infrastructure adjustments) with 4% discount rate ($\pi = 0.04$) which is the common value used for public investments in the Netherlands. Also, total travel cost for the peak hour is converted to a yearly basis. Peak hour travel cost is assumed to be 1/8 of 24-hr travel cost ($\lambda = 8 * 30 * 12$), therefore ($\sigma = \lambda(1{+}\pi)^{(\tau\text{-}1)} = 5{,}945$). PS correction parameter value is set to 1 for this case. Logit route choice parameter values for CVs and AVs are respectively 1.25 and 2. The BPR function parameters (α & b) used are 0.15 and 4, respectively. A comprehensive description of the network, demand, and optimal results obtained by each algorithm is provided in Table 2.

A full enumeration of all possible solutions is also performed to find the global optimum of this problem for the 50% penetration rate scenario. Note that this is only feasible for very small networks. Figure 4 exhibits the graph for the global optimum point of the problem and Table 2 reports network performance criteria related to this point.

All three algorithms easily find the global optimum point of the problem (Figure 4) on every run in less than a minute with an average common computer without serious parameter tuning. This confirms that these algorithms are capable of finding global optimum solutions of small size problems. Since this problem is trivial, parameter tuning and computational issues are not relevant here. These will be discussed in the next case study.

5.2 | Case study 2: Delft network

In this section, we use another case study on a more complex network to compare the performance of different algorithms and to discuss relevant practical implications. A network similar to the road network of Delft, the Netherlands is used and the network data and the demand patterns are based on a peak hour in a tutorial project in OmniTRANS transport modeling software package. It includes 1,151 links,

TABLE 2 Description of the demonstration case with a synthetic network (demand is 280 v/hr between each pair of nodes 1–9)

Network description					
Link pairs XY (node X to node Y)	**Capacity (v/hr)**	**Length (km)**	**Speed (km/hr)**	**Road type**	**Adjustment cost (€/km)**
12–21	500	3	40	Local road	Infeasible
23–32	500	3	40	Local road	Infeasible
36–63	4,000	3	120	Motorway	300,000
69–96	4,000	3	120	Motorway	300,000
98–89	500	3	40	Local road	Infeasible
87–78	500	3	40	Local road	Infeasible
74–47	4,000	3	120	Motorway	300,000
41–14	4,000	3	120	Motorway	300,000
45–54	2,000	3	80	Expressway	1,200,000
56–65	2,000	3	80	Expressway	1,200,000
25–52	2,000	3	80	Expressway	1,200,000
58–85	2,000	3	80	Expressway	1,200,000
Optimal results					
Objective function	**TTC**	**TTT**	**TTD**	**TAC**	**Connected**
48,440	44,807	2,803	130,559	21,600,000	Yes

494 nodes, and 462 origin-destination pairs. A subset of links corresponding to motorways, expressways, and main roads are selected as feasible links for AD subnetwork. Analysis of the AV autopilot disengagement reports mentioned in the Introduction confirms that these are the most suitable roads for AVs. We consider different adjustment costs per road type which are proportional to the road hierarchy shown in Figure 5a. The reason is that we want roads in AD subnetwork to meet certain design and quality standards. Motorways usually have highest standards and can be suitable for AD with minimal adjustments. Freeways and urban arterials can have high standards as well but they might include segments with unsegregated traffic and complex intersections that require adjustments given that 89% of the reported accidents involving an AV have happened at an intersection (Favarò et al., 2017). Since we are not aware of any study that provides accurate cost estimates for the adjustments mentioned in the introduction, we assumed the following costs and provide a sensitivity analysis to show how variations in the adjustment costs can affect the results: 100,000 €/km for local roads, 30,000 for freeways, and 5,000 for motorways. The total number of feasible links is 421. We considered three demand scenarios with 10%, 50%, and 90% penetration rate of AVs as well as a base case scenario with no AVs. Furthermore, a variation where all feasible links are included in the AD subnetwork is applied for each demand scenario to provide lower bounds for comparisons. The value of σ is the same as mentioned in the previous subsection.

Different road types, feasible roads for AD subnetwork, and congestion patterns in the base case of Delft network are shown in Figure 5. Table 3 summarizes other important parameters used in this case study. ULMP parameters are determined after extensive trials to optimize each algorithm's performance. LLMP parameters are the same as those used in Madadi et al. (2019) or within the sensitivity analysis bandwidth used in that study.

5.3 | Numerical results and analysis

Since there is stochasticity in the algorithms, we used five independent runs for each scenario and reported the average results. For the objective function, a 95% confidence interval is reported to show the stability of the algorithms. The optimal graphs and convergence curves reported are from the runs with best results. In the following subsections, we compare the results based on each of three mentioned criteria.

5.3.1 | Effectiveness

Since with heuristic algorithms, there is no guarantee for finding the global optimum, we consider the degree of

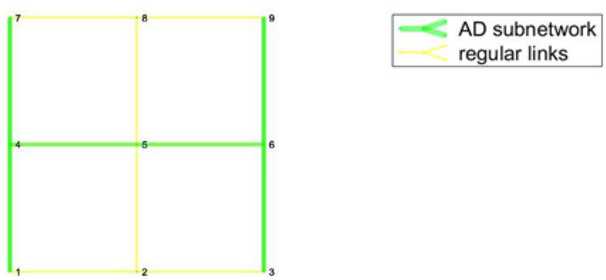
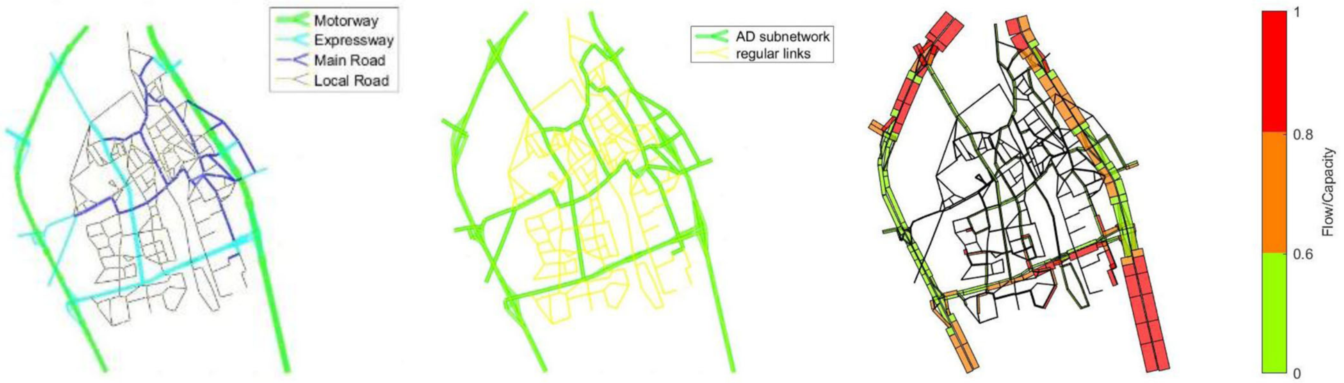


FIGURE 4 Optimal AD subnetwork for synthetic network and 50% penetration rate scenario



(a) Road types

(b) Feasible roads for AD subnetwork

(c) Congestion pattern (bandwidth represents relative flow)

FIGURE 5 (a) Road types, (b) feasible roads for AD subnetwork, and (c) congestion patterns in Delft case study (base case)

TABLE 3 Important input parameters used in Delft case study

Upper level problem (ULMP) parameters

Algorithm	Parameter	Value
ELS	Population size (n^p)	10
	Merging interval (i_m)	20
	Number of candidates (n_c^p)	4
GA	Population size (n^p)	100
	Elite size (n_e^p)	20
	Maximum generations (n^g)	150
	Crossover fraction (c_f)	0.8
MGA	Population size (n^p)	300
	Elite size (n_e^p)	30
	Maximum generations (n^g)	200
	Crossover fraction (c_f)	0.8
	Penalty weight (w_κ)	2,000

Lower level problem (LLMP) parameters

		Penetration rate	
Parameter	**Class**	**10%**	**[50–90%]**
PCU	CV	1	1
	AV	0.95	0.9
VoTT (€/hr)	CV	9	9
	AV	7.2	7.2
VoTD (€/km)	CV	0.19	0.19
	AV	0.114	0.114

optimality (i.e., how low the value of the objective function is) as the main indicator for effectiveness of an algorithm. As demonstrated in Table 4, in terms of degree of optimality, ELS narrowly outperforms GA and MGA in all scenarios in this problem. Although, both GA and MGA (especially MGA), are competitive in terms of objective function value in all scenarios. Furthermore, the confidence intervals reported overlap in most cases but ELS shows more stability within different runs. An interesting observation regarding TTC is that MGA outperforms the other two in terms of TTC in all scenarios, yet this is achieved at price of significantly higher TAC that leads to less desirable objective function values. Regarding total travel time (TTT), the differences in most cases are not significant. This can be explained partially by the fact that a portion of gains in TTC are due to the assumption of lower VoTD and VoTT for AVs. Moreover, congestion reduction gains of AD are, to some extent, counterbalanced by longer detours for routes using the AD subnetwork. Trends in total travel distance (TTD) corroborate this explanation since less effective results in terms of TTC have lower TTD. That is, more effective designs cause more rerouting towards the AD subnetwork which lead to higher travel distances.

Another discernible pattern is related to the stretches of roads included in the best results. It can be seen in Figures 6–8 that the subnetworks obtained by all three algorithms, include almost identical main stretches of roads in all scenarios. This suggests plausibility for the designs acquired by these algorithms. Given that the TTT and TTC values are, in most cases, very similar to the ones in "all feasible links selected" variation that has been used as a lower bound for TTC and TTT.

5.3.2 | Design quality

Since an AD subnetwork is primarily designed to serve the flow of AVs in automated mode with minimum possible disruptions, connectivity is a crucial factor for its quality and effectiveness. Yet a strong connectivity is unnecessary given that AD subnetworks are designed to facilitate mobility rather than accessibility. There is no need to reach every single node from every other node in both directions within the AD subnetwork. A weak connectivity (as defined in Definition 3.3.5) is sufficient to ensure that if a part of a trip is possible within the AD subnetwork, there is no interruption (i.e., switching between driving modes)

TABLE 4 Comparison of algorithm performances in effectiveness (optimality)

Algorithm	Objective function	TTC	TTT	TTD	TAC
10% penetration rate					
ELS	100,982 ± 0.63	100,749	5,508	279,008	1,380,488
GA	100,998 ± 8.87	100,753	5,508	279,001	1,461,647
MGA	101,052 ± 6.16	100,708	5,508	279,006	2,047,384
All feasible links selected	101,449	100,561	5,508	279,006	5,281,545
"As is"	102,519	102,519	5,509	278,628	0
50% penetration rate					
Algorithm	**Objective function**	**TTC**	**TTT**	**TTD**	**TAC**
ELS	93,371 ± 0.84	92,798	5,492	280,528	3,407,849
GA	93,437 ± 8.11	92,846	5,492	280,524	3,509,915
MGA	93,382 ± 5.01	92,790	5,492	280,527	3,519,558
All feasible links selected	93,546	92,658	5,492	280,525	5,281,545
"As is"	102,519	102,519	5,509	278,628	0
90% penetration rate					
Algorithm	**Objective function**	**TTC**	**TTT**	**TTD**	**TAC**
ELS	85,673 ± 0.93	84,969	5,491	282,039	4,182,494
GA	85,783 ± 17.9	85,107	5,491	282,039	4,019,448
MGA	85,678 ± 4.11	84,962	5,491	282,039	4,257,011
All feasible links selected	85,789	84,901	5,491	282,038	5,281,545
"As is"	102,519	102,519	5,509	278,628	0

Note: Reported values are averages of five independent runs. ± signs denote standard errors for 95% confidence intervals.

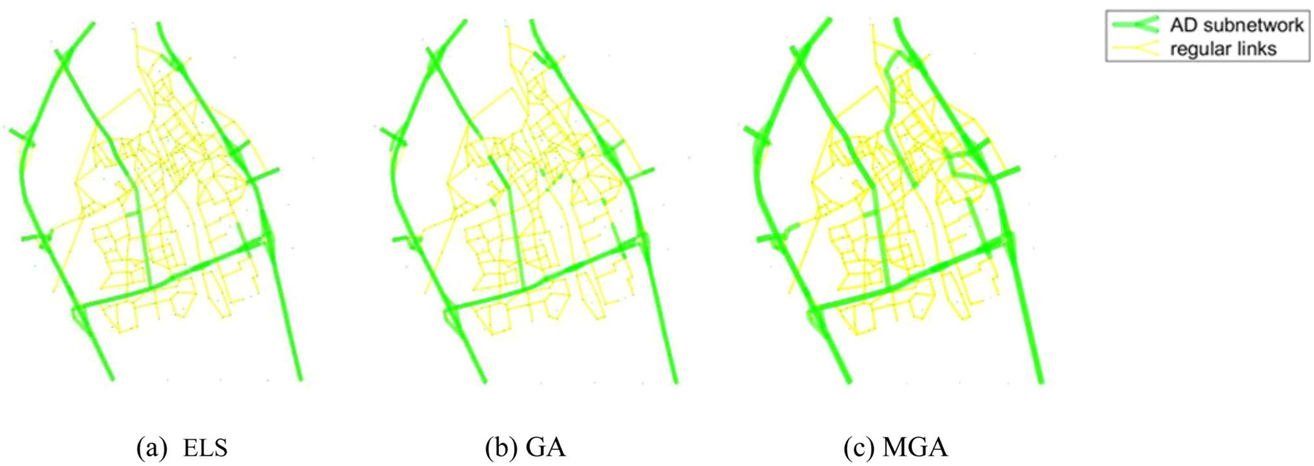


FIGURE 6 Optimal AD subnetworks (best results) for 10% penetration scenario: (a) ELS, (b) GA, and (c) MGA

in that part of the trip. Therefore, we considered connectivity as the criterion for comparing the design quality of the results.

It is evident from Figures 6–8 that all the optimal designs produced with ELS and MGA meet the connectivity criterion as defined in this paper but most of the designs produced by GA do not meet this criterion. This was to be expected since GA does not include any mechanism to favor connected designs. Although, apart from a limited number of isolated links and stretches, GA has produced very similar designs compared to those produced by ELS. Also, all designs generated by MGA meet the connectivity criterion suggesting that the penalty function (Equation (14)) has been effective. However, this has resulted in substantial additional computation times for MGA. Efficiency of the algorithms will be discussed in the next subsection.

5.3.3 | Efficiency and convergence

Table 5 summarizes the computation times for all algorithms in all scenarios. ELS has the lowest computation time in 10%

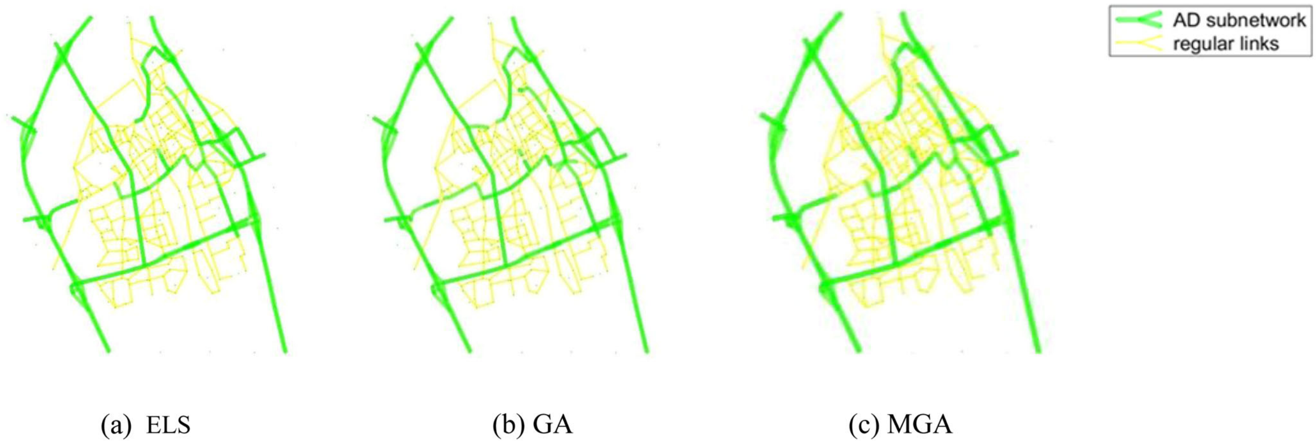


FIGURE 7 Optimal AD subnetworks (best results) for 50% penetration scenario: (a) ELS, (b) GA, and (c) MGA

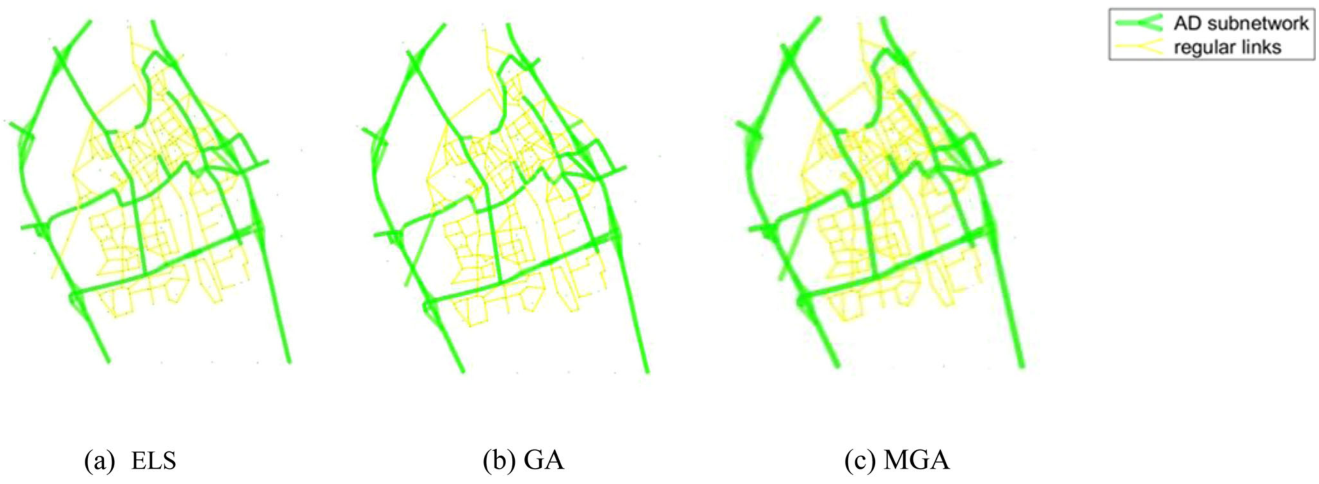


FIGURE 8 Optimal AD subnetworks (best results) for 90% penetration scenario: (a) ELS, (b) GA, and (c) MGA

TABLE 5 Average computation times in Delft case study

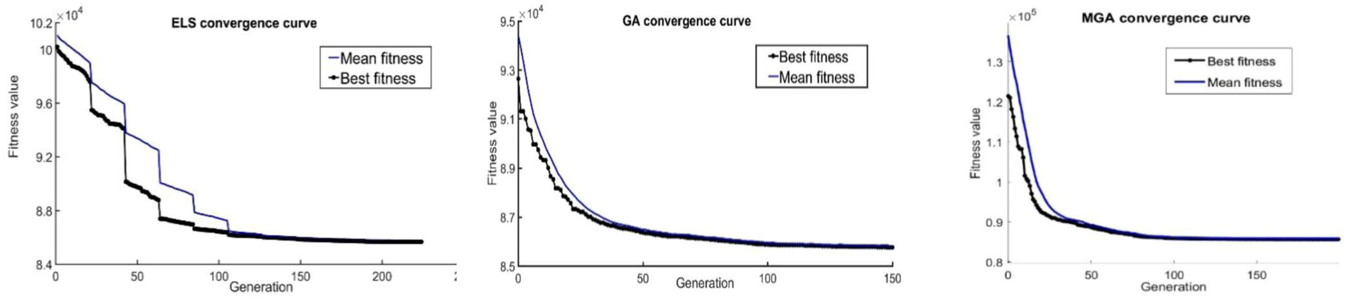
	Average computation time (minutes)		
Algorithm	**10%**	**50%**	**90%**
ELS	76	140	157
GA	271	274	270
MGA	964	968	966

penetration rate scenario but this value increases with the increase in penetration rate. The reason is that ELS extends one link at a time, so when the optimal AD subnetwork includes more links (with more AVs) computation times increase, yet this increase seems to be linear. GA shows stable computation times for different scenarios but they are higher compared to ELS in all scenarios. Finally, MGA shows stable computation times cross scenarios as well. Although, its numbers are significantly higher compared to all others (more than three times the computation times of GA). This confirms that the penalty function, although effective in securing connected designs, is rather inefficient. This can become a major issue when applying MGA on large scale networks, especially since computation times tend to increase exponentially with the number of links.

Regarding the ULMP convergence curves presented in Figure 9, GA and MGA display a common evolutionary behavior with the average population following the elite with a lag and algorithms settling down toward the end.

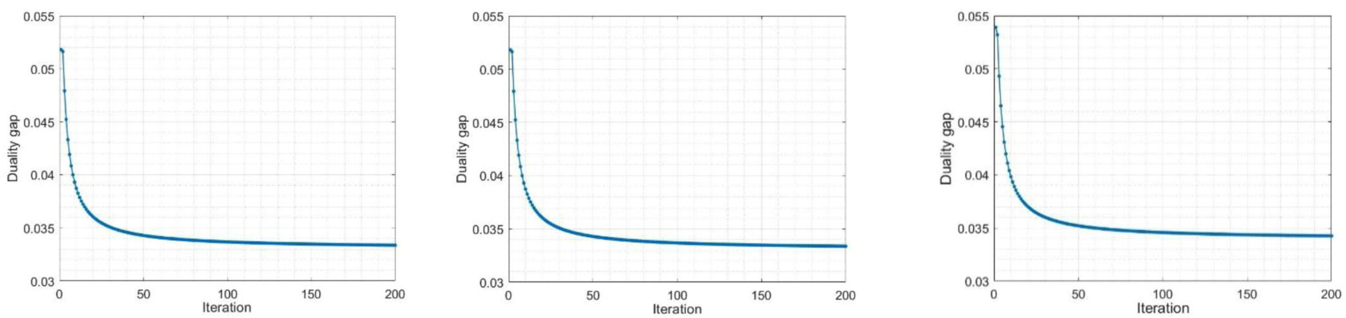
The same pattern is noticeable for ELS with the addition of sudden significant improvements in fitness values that correspond to merging intervals (refer to Table 3). This implies that the merging operation (step 3 of ELS steps) substantially accelerates the convergence of ELS. Regarding MGA, two stages of improvements are noticeable from the graph. The first stage which includes fast and substantial improvements in the beginning, shows how MGA is attempting to find designs with less connected components to reduce the penalty. The second stage denotes the search for more effective designs.

Since ULMP is evaluated when LLMP is at its equilibrium, the convergence curves (duality gaps) of LLMP for



(a) ELS (b) GA (c) MGA

FIGURE 9 Convergence curves (ULMP) in 90% penetration scenario: (a) ELS, (b) GA, and (c) MGA



(a) ELS (b) GA (c) MGA

FIGURE 10 Convergence curves of the LLMP (duality gaps) for optimal results in 50% penetration scenario: (a) ELS, (b) GA, and (c) MGA

TABLE 6 Sensitivity analysis (variations in results) with respect to demand and adjustment cost variations in 50% penetration rate scenario (obtained by ELS)

Demand variation	Objective function	TTC	TTT	TTD	TAC
+20%	+31%	+31%	+41%	+20%	+7%
+10%	+14%	+14%	+18%	+10%	+5%
−10%	−13%	−13%	−15%	−10%	−3%
−20%	−24%	−24%	−28%	−20%	−20%
Adjustment cost variation					
+300%	+1.21%	+0.68%	+0.02%	0.00%	+87%
+100%	+0.55%	+0.33%	+0.02%	0.00%	+36%
−25%	−0.16%	−0.07%	−0.00%	0.00%	−16%
−50%	−0.35%	−0.12%	−0.00%	0.00%	−37%

optimal results of the 50% penetration rate scenario are shown in Figure 10 as an example. A reasonable duality gap is reached (even within the first few iterations) considering that LLMP is a SUE and that the duality gap for a SUE should not approach zero. For the intermediate iterations of ULMP, the solution of LLMP terminates when a threshold of 4% duality gap is reached which is very close to the lowest duality gap achieved after 200 iterations (3.3%). Also, given that all algorithms in all their fitness function evaluations use the same threshold, the comparison is fair.

5.3.4 | Sensitivity analysis

An average peak hour traffic demand is used for the case studies here (as is the case in most NDPs). However, day-to-day variations in demand are common in transportation networks. Therefore a sensitivity analysis on demand is reported in Table 6 for 50% penetration rate scenario to show how variations in demand can cause variations in objectives. As it can be seen in Table 6, increases in demand and congestion levels have more impact on TTC, TTT, and objective function compared to decreases. TTDs change proportional to demand and TACs show more sensitivity to reductions in demand after a certain point. One explanation is that with less demand, less investment is justified, yet significant increases in demand cause too much congestion that cannot be mitigated with extra investments. Therefore, proper demand estimation is crucial for practical applications.

Moreover, sensitivity analysis for variations in adjustment costs reveal that these changes have minor impacts on network performance indicators. However, these variations affect the optimal configuration of the AD subnetwork, as evidenced by the changes in TACs whose sensitivities change with different proportions with respect to increases and decreases in the adjustment costs.

6 | DISCUSSION AND CONCLUSIONS

In this paper, we considered the problem of optimizing urban road networks for a mixture of AVs and CVs. We modeled this problem as a bi-level DNDP and attempted to optimize the trade-offs between the costs and the benefits of a certain network configuration, namely the AD subnetwork. Moreover, we suggested a solution algorithm for this problem and benchmarked its performance against two solution algorithms for DNDP using a case study and considering three different performance criteria. The results reveal that the ELS algorithm presented in this study has a satisfactory performance in all three criteria considered, namely, effectiveness, efficiency, and design quality.

Design quality criteria is specifically relevant to the formulation of NDP introduced in this study, which enforces the connectivity constraint on the subnetworks. It is especially this constraint that makes the commonly used NDP solutions less suitable, while the MGA (i.e., the GA including a penalty for disconnected subnetworks) proves to be inefficient. The ELS algorithm developed in this study is better tailored to this new formulation of the NDP and proves to be efficient in dealing with connectivity constraints.

Moreover, our findings indicate that the optimal layout of AD subnetworks depend on demand and adjustment cost; lower penetration rates of AVs call for less dense subnetworks and less investment. Nonetheless, a large proportion of all possible network performance benefits are achievable with low-density AD subnetworks. This is, to some extent, due to the fact that more effective designs usually include roads with highest capacity first. Therefore, it is logical to redirect AD to designated parts of the network that are selected based on their inherent safety and optimized for network performance. On the other hand, higher penetration rates of AVs demand denser AD subnetworks and can deliver more benefits. A key observation was that an effective subnetwork can deliver a large proportion of the benefits obtained by upgrading all feasible links (which includes substantial costs), especially with higher penetration rates of AVs. Constructing AD subnetworks has apparent safety and performance benefits since it includes redirecting AV traffic to specific parts of the network that are either naturally safer or have been adjusted to be safer, and are optimized for network performance. Therefore, it is recommendable for urban planners and transport authorities to consider such configurations for the transition period.

Demand for AVs will evolve gradually over time. So AD subnetworks should be extended gradually as well. It is practical to start investing on high-impact places first and gradually extending the AV-friendly part of the network with the increase in the market penetration rate of AVs to maintain the optimal trade-off between the investments and the benefits over time.

Therefore, a possible future research direction is considering the time dimension and modeling the problem as a tri-level NDP with timing as the third level and including AV demand elasticity. This can provide an optimal plan for gradually upgrading the AD subnetwork over a time horizon in reaction to AV demand.

Another possible extension of this problem is considering DTA models for the lower level equilibrium problem to model the behavioral differences of CVs and AVs more accurately. This can also allow for elaborate modeling of intersections, which are important elements in urban road networks, in presence of CVs and AVs. However, this can be extremely challenging due to the high computation requirements of DTA models and the fact that the literature on DTA models for AVs is still scarce.

ACKNOWLEDGMENTS

This study was funded as part of the project Spatial and Transport Impacts of Automated Driving (STAD) by the Netherlands Organisation for Scientific Research (NWO) under contract 438-15-161. We also would like to thank the Editor and the seven anonymous reviewers for their constructive comments and valuable suggestions to improve the quality of the article.

REFERENCES

Chakroborty, P. (2003). Genetic algorithms for optimal urban transit network design. *Computer-Aided Civil and Infrastructure Engineering*, *18*, 184–200. https://doi.org/10.1111/1467-8667.00309

Chen, Z., He, F., Yin, Y., & Du, Y. (2017). Optimal design of autonomous vehicle zones in transportation networks. *Transportation Research Part B: Methodological*, *99*, 44–61. https://doi.org/10.1016/j.trb.2016.12.021

Chen, Z., He, F., Zhang, L., & Yin, Y. (2016). Optimal deployment of autonomous vehicle lanes with endogenous market penetration. *Transportation Research Part C: Emerging Technologies*, *72*, 143–156. https://doi.org/10.1016/j.trc.2016.09.013

Correia, G., de Looff, E., van Cranenburgh, S., Snelder, M., & van Arem, B. (2019). On the impact of vehicle automation on the value of travel time while performing work and leisure activities in a car: Theoretical insights and results from a stated preference survey. *Transportation Research Part A*, *119*, 359–382. https://doi.org/10.1016/j.tra.2018.11.016

Courbon, T., Scharningg, K., Techner, T., Innamaa, S., Kulmala, R., Alkim, T., … Rausch, J. (2016). *EU-EIP Activity 4.2 Facilitating automated driving, Task 1: Identification of requirements towards network operators* (Technical report). The Hague, The Netherlands: European ITS Platform.

de Looff, E., Correia, G. H. de A., van Cranenburgh, S., Snelder, M., & van Arem, B. (2018, January). *Potential changes in value of travel time as a result of vehicle automation : A case-study in the Netherlands*. Paper presented at the 97th annual meeting of Transportation Research Board, Washington, DC.

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: Disengagements, accidents and reaction times. *PLoS ONE*, *11*, 1–14. https://doi.org/10.1371/journal.pone.0168054

Even, S., & Tarjan, R. (1975). Network flow and testing graph connectivity. *SIAM Journal on Computing*, *4*, 507–518.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Farah, H., Erkens, S. M. J. G., Alkim, T., & van Arem, B. (2018). Infrastructure for automated and connected driving: State of the art and future research directions. In G. Meyer & S. Beiker (Eds.), *Road vehicle automation 4* (pp. 187–197). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-319-60934-8

Farahani, R. Z., & Miandoabchi, E. (2012). *Graph theory for operations research and management: Applications in industrial engineering*. Hershey, PA: IGI Global.

Farahani, R. Z., Miandoabchi, E., Szeto, W. Y., & Rashidi, H. (2013). A review of urban transportation network design problems. *European Journal of Operational Research*, *229*, 281–302. https://doi.org/10.1016/j.ejor.2013.01.001

Favarò, F., Eurich, S., & Nader, N. (2018). Autonomous vehicles' disengagements: Trends, triggers, and regulatory limitations. *Accident Analysis and Prevention*, *110*, 136–148. https://doi.org/10.1016/j.aap.2017.11.001

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M., & Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California. *PLoS ONE*, *12*, 1–20. https://doi.org/10.1371/journal.pone.0184952

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Boston, MA: Addison-Wesley Longman Publishing Co., Inc.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: University of Michigan Press.

Levin, M. W., & Boyles, S. D. (2015). Effects of autonomous vehicle ownership on trip, mode and route choice. *Transportation Research Record: Journal of the Transportation Research Board*, *2493*, 29–38. https://doi.org/10.3141/2493-04

Levin, M. W., & Boyles, S. D. (2016). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, *62*, 103–116. https://doi.org/10.1016/j.trc.2015.10.005

Le Vine, S., Zolfaghari, A., & Polak, J. (2015). Autonomous cars: The tension between occupant experience and intersection capacity. *Transportation Research Part C: Emerging Technologies*, *52*, 1–14. https://doi.org/10.1016/j.trc.2015.01.002

Lv, C., Cao, D., Zhao, Y., Auger, D. J., Sullman, M., Wang, H., … Mouzakitis, A. (2018). Analysis of autopilot disengagements occurring during autonomous vehicle testing. *IEEE/CAA Journal of Automatica Sinica*, *5*, 58–68. https://doi.org/10.1109/JAS.2017.7510745

Madadi, B., van Nes, R., Snelder, M., & van Arem, B. (2019). Assessing the travel impacts of subnetworks for automated driving: An exploratory study. *Case Studies on Transport Policy*, *7*, 48–56. https://doi.org/10.1016/j.cstp.2018.11.006

Magnanti, T. L., & Wong, R. T. (1984). Network design and transportation planning: Models and algorithms. *Transportation Science*, *18*(1), 55.

Mahmassani, H. S. (2016). 50th anniversary invited article—Autonomous vehicles and connected vehicle systems: Flow and operations considerations. *Transportation Science*, *50*, 1140–1162.

Migdalas, A. (1995). Bilevel programming in traffic planning: Models, methods and challenge. *Journal of Global Optimization*, *7*, 381–405. https://doi.org/10.1007/BF01099649

Milakis, D., Snelder, M., van Arem, B., van Wee, B., & Correia, G. H. de A. (2017). Development and transport implications of automated vehicles in the Netherlands: Scenarios for 2030 and 2050. *European Journal of Transport and Infrastructure Research*, *17*, 63–85.

Milakis, D., van Arem, B., & van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems*, *21*, 324–348. https://doi.org/10.1080/15472450.2017.1291351

Nitsche, P., Mocanu, I., & Reinthaler, M. (2014, November). *Requirements on tomorrow's road infrastructure for highly automated driving*. Paper presented at the IEEE International Conference on Connected Vehicles and Expo, Vienna, Austria. https://doi.org/10.1109/ICCVE.2014.7297694

Nowakowski, C., O'Connell, J., Shladover, S. E., & Cody, D. (2010, September). *Cooperative adaptive cruise control: Driver acceptance of following gap settings less than one second*. Paper presented at the 54th annual meeting of the Human Factors and Ergonomics Society, San Francisco, CA. https://doi.org/10.1177/154193121005402403

Patel, R., Levin, M. W., & Boyles, S. D. (2016). Effects of autonomous vehicle behavior on arterial and freeway networks. *Transportation Research Record: Journal of the Transportation Research Board*, *2561*, 9–17. https://doi.org/10.3141/2561-02

Poorzahedy, H., & Rouhani, O. M. (2007). Hybrid meta-heuristic algorithms for solving network design problem. *European Journal of Operational Research*, *182*, 578–596. https://doi.org/10.1016/j.ejor.2006.07.038

Puylaert, S., Snelder, M., van Nes, R., & van Arem, B. (2018). Mobility impacts of early forms of automated driving—A system dynamic approach. *Transport Policy*, *72*, 171–179. https://doi.org/https://doi.org/10.1016/j.tranpol.2018.02.013

Rios-Torres, J., & Malikopoulos, A. A. (2017). Automated and cooperative vehicle merging at highway on-ramps. *IEEE Transactions on Intelligent Transportation Systems*, *18*, 780–789. https://doi.org/10.1109/TITS.2016.2587582

SAE International. (2018). *Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles*. Warrendale, PA: SAE International.

Shida, M., & Nemoto, Y. (2009, September). *Development of a small-distance vehicle platooning system*. Paper presented at the 16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services, Stockholm, Sweden.

Shladover, S., Su, D., & Lu, X-Y. (2012). Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transportation Research Record: Journal of the Transportation Research Board*, *2324*(1), 63–70.

Shladover, S. E. (2016). The truth about "self-driving" cars. *Scientific American*, *314*, 52–57. https://doi.org/10.1038/scientificamerican0616-52

Shladover, S. E. (2018). Connected and automated vehicle systems: Introduction and overview. *Journal of Intelligent Transportation Systems*, *22*(3), 190–200. https://doi.org/10.1080/15472450.2017.1336053

Shladover, S. E., Nowakowski, C., Lu, X.-Y., & Ferlis, R. (2015). Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record: Journal of the Transportation Research Board*, *2489*, 145–152. https://doi.org/10.3141/2489-17

van Arem, B., van Driel, C. J. G., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on Intelligent Transportation Systems*, *7*, 429–436. https://doi.org/10.1109/TITS.2006.884615

Wegman, F., & Aarts, L. (Ed.). (2006). *Advancing sustainable safety: National road safety outlook for 2005-2020*. The Netherlands: SWOV Institute for Road Safety Research. https://doi.org/10.1016/j.ssci.2007.06.013

Wegman, F., Aarts, L., & Bax, C. (2008). Advancing sustainable safety: National road safety outlook for The Netherlands for 2005–2020. *Safety Science*, *46*, 323–343. https://doi.org/10.1016/j.ssci.2007.06.013

Wu, J. H., Florian, M., & He, S. (2006). An algorithm for multi-class network equilibrium problem in PCE of trucks: Application to the SCAG travel demand model. *Transportmetrica*, *2*, 1–9. https://doi.org/10.1080/18128600608685656

Yang, H., & Bell, M. (1998). Models and algorithms for road network design: A review and some new developments. *Transport Reviews*, *18*, 257–278. https://doi.org/10.1080/01441649808717016

Ye, Y., & Wang, H. (2018). Optimal design of transportation networks with automated vehicle links and congestion pricing. *Journal of Advanced Transportation*, *2018*, 1–12.

Zhang, Y. (2013). *Adapting infrastructure for automated driving* (Technical report). Tampa, FL: Center for Urban Transportation Research (CUTR). https://docplayer.net/38423350-Adapting-infrastructure-for-automated-driving.html

How to cite this article: Madadi B, van Nes R, Snelder M, van Arem B. A bi-level model to optimize road networks for a mixture of manual and automated driving: An evolutionary local search algorithm. *Comput Aided Civ Inf*. 2019;1–17. https://doi.org/10.1111/mice.12498