applied sciences

MDPI

Article

Optimal Lane Allocation Strategy in Toll Stations for Mixed Human-Driven and Autonomous Vehicles

Zuoyu Chai [1,2], Tanghong Ran [1] and Min Xu [1,*]

[1] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China; zy_chai@seu.edu.cn (Z.C.); tanghong.ran@connect.polyu.hk (T.R.)
[2] School of Transportation, Southeast University, Nanjing 211189, China
* Correspondence: min.m.xu@polyu.edu.hk; Tel.: +852-2766-6593

Abstract: Highway toll stations are equipped with electronic toll collection (ETC) lanes and manual toll collection (MTC) lanes. It is anticipated that connected autonomous vehicles (CAVs), MTC human-driven vehicles (MTC-HVs), and ETC human-driven vehicles (ETC-HVs) will coexist for a long time, sharing toll station infrastructure. To fully leverage the congestion reduction potential of ETC, this paper addresses the problem of ETC lane allocation at toll stations under heterogeneous traffic flows, modeling it as a mixed-integer nonlinear bilevel programming problem (MINLBP). The objective is to minimize total toll station travel time by optimizing the number of ETC lanes at station entrances and exits while considering ETC-HVs' lane selection behavior based on the user equilibrium principle. As both upper-level and lower-level problems are convex, the bilevel problem is transformed into an equivalent single-level optimization using the Karush–Kuhn–Tucker (KKT) conditions of the lower-level problem, and numerical solutions are obtained using the commercial solver Gurobi. Based on surveillance video data from the Liulin toll station (Lianhuo Expressway) in Zhengzhou, China, numerical experiments were conducted. The results illustrate that the proposed method reduces total vehicle travel time by 90.44% compared to the current lane allocation scheme or the proportional lane allocation method. Increasing the proportion of CAVs or ETC-HVs helps accommodate high traffic demand. Dynamically adjusting lane allocation in response to variations in traffic arrival rates is proven to be a more effective supply strategy than static allocation. Moreover, regarding the interesting conclusion that all ETC-HVs choose the ETC lanes, we derived the relaxed analytical solution of MINLBP using a parameter iteration method. The analytical solution confirmed the validity of the numerical experiment results. The findings of this study can effectively and conveniently guide lane allocation at highway toll stations to improve traffic efficiency.

Keywords: heterogeneous traffic; queueing theory; user equilibrium; MINLBP; lane allocation; transportation planning and management

Academic Editor: Yutaka Ishibashi

Received: 16 November 2024
Revised: 23 December 2024
Accepted: 1 January 2025
Published: 2 January 2025

Citation: Chai, Z.; Ran, T.; Xu, M. Optimal Lane Allocation Strategy in Toll Stations for Mixed Human-Driven and Autonomous Vehicles. *Appl. Sci.* **2025**, *15*, 364. https://doi.org/10.3390/app15010364

1. Introduction

With the rapid development of the economy and the continuous increase in the total length of highways and the number of vehicles, the demand for highways is rising. However, the rapid growth of vehicle numbers does not match the speed of infrastructure development, leading to increasingly prominent traffic congestion issues [1]. Promoting a balance between the supply and demand of highways can no longer rely solely on increasing supply [2]. Instead, it should focus on optimizing the efficiency of existing resources from the perspective of traffic management and control.

Toll stations are an essential component of highways, and as intermittent flow facilities, they are essentially bottlenecks in traffic flow [3,4]. If the toll station's capacity is insufficient, vehicles queuing and congesting at the toll station will limit the overall traffic efficiency of the highway. Congestion on highways and toll booths will have a huge cost [3,5,6]. Therefore, improving the toll station's capacity is one of the key measures to ensure the efficient operation of highways. Recognizing the severity of this issue, an increasing number of countries, such as China [7,8], the United States [9], India [10], Japan [11], and Serbia [12,13], are devoting greater attention to addressing congestion and queuing problems at highway toll stations.

Currently, the toll collection methods on highways are diversifying, including Electronic Toll Collection (ETC for short) and Manual Toll Collection (MTC for short) [14,15]. However, if the distribution of ETC lanes and MTC lanes is not properly balanced, it may result in queuing congestion at MTC lanes, while ETC lanes remain idle for extended periods. In addition to the diversification of toll methods, the categories of highway users are also becoming more varied [16]. As shown in Figure 1, based on different tolling methods, users can be roughly classified into three categories. First, Connected Autonomous Vehicles (CAVs) can only use ETC lanes for automatic toll payment throughout the journey. Second, manual toll-collection human-driven vehicles (MTC-HVs) must use MTC lanes for payment. Finally, electronic toll collection human-driven vehicles (ETC-HVs) can flexibly choose between ETC or MTC lanes. This diversification of user types reflects the trend in transportation technology development and also places higher demands on the management of toll systems. It requires further optimization of lane allocation and management strategies to accommodate the needs of different types of vehicles. Moreover, vehicle composition and traffic demands vary across different countries and regions at different times, and this dynamic demand requires toll systems to have sufficient flexibility to adapt to traffic fluctuations [12,17].

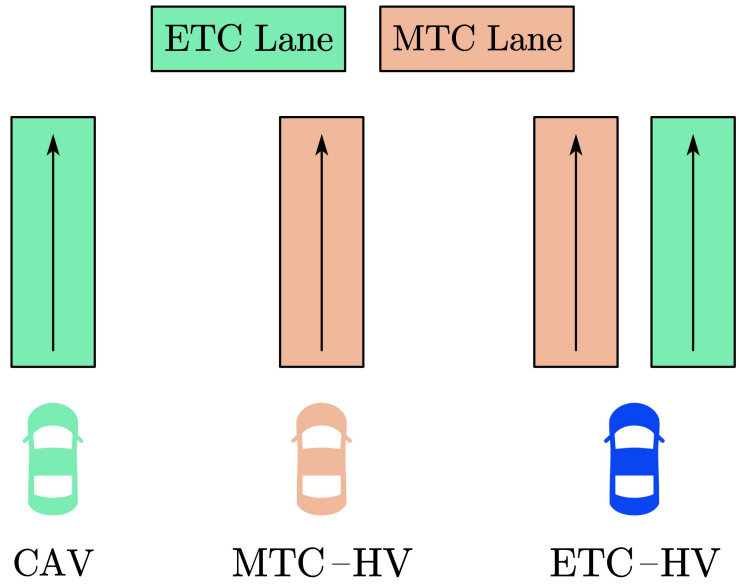

Figure 1. The selectable lanes for different types of vehicles.

Therefore, the rational distribution of ETC and MTC lanes at toll stations is an effective way to improve highway traffic efficiency [18].

1.1. Literature Review

Under heterogeneous traffic flow conditions, significant progress has been made in research on toll station lane allocation. Researchers have analyzed the traffic characteristics

of different vehicle types, driver decision preferences, and behavioral patterns, uncovering key insights into how various vehicles navigate toll stations within diverse traffic flows [17,19–21]. Specifically, studies have developed driver decision models to predict lane choices, examined behavioral changes under congestion to optimize lane allocation strategies, and introduced intelligent guidance technologies to enhance passage efficiency. These findings have not only deepened the understanding of driver behavior but also provided a foundation for devising rational lane management strategies, effectively alleviating toll station congestion and improving overall traffic system efficiency and safety.

In toll lane optimization research based on queuing theory, researchers have developed allocation models for multi-lane and multi-vehicle scenarios [13,22]. Queuing theory, particularly the classical $M/M/1$ [8,14] model and its extensions (e.g., $M/M/N$ [23,24], $M/G/1$ [25] and $M/G/N$ [22,26] models), have been widely applied in traffic engineering and toll plaza analysis, such as capacity planning, real-time traffic control strategies, automated tolling system evaluation, and so on. The $M/M/1$ model has the advantages of simplicity in computation, intuitive analysis, and ease of application. It is particularly effective when vehicle arrival intervals and service times follow exponential distributions, allowing for the quick calculation of system performance metrics such as average waiting time and queue length. Despite its simplicity and frequent use, traffic flows can sometimes be more complex, with vehicle arrival intervals and service times not necessarily following exponential distributions. The $M/M/N$ model extends the $M/M/1$ model by considering systems with multiple parallel servers (lanes) to handle vehicle traffic. Sometimes users are unable to switch service lanes after entering the queue system, forming N independent $M/M/1$ queuing systems. Similarly, the $M/G/N$ model generalizes the $M/M/N$ model by allowing both arrival intervals and service times to follow general distributions. This model is suited for scenarios where the assumption of exponential distribution no longer holds for either arrival or service processes. These models provide a solid theoretical foundation for more accurate simulations of toll station traffic.

Additionally, with the integration of intelligent optimization algorithms (e.g., genetic algorithms, simulated annealing, and particle swarm optimization), these models can efficiently converge to optimal or near-optimal solutions under multi-variable and multi-constraint conditions, significantly improving passage efficiency [8,27]. By combining dynamic traffic flow theory with real-time data, researchers have also developed models capable of dynamically adjusting lane allocation strategies to adapt to changing traffic volumes, thereby further reducing congestion. This body of research not only supports the efficient utilization of toll station resources but also demonstrates substantial potential for designing intelligent traffic management systems.

With advancements in predictive analytics and machine learning technologies, toll station management has progressively shifted towards a data-driven paradigm. By modeling and analyzing historical and real-time data, predictive analytics accurately forecasts traffic volume and congestion, offering proactive guidance for flow management [4,28,29]. Concurrently, machine learning refines dynamic scheduling and resource allocation models by extracting complex patterns from large-scale data, significantly enhancing efficiency [12]. Recent breakthroughs focus on deep learning-based traffic forecasting, adaptive toll pricing strategies, and reinforcement learning applications, enabling toll stations to optimize automatically based on real-time feedback [13,17,22]. These advancements have not only boosted toll station efficiency and service quality but also laid a foundation for the development of intelligent traffic systems. Table 1 lists some of the existing related work.

Table 1. Existing related work.

Paper	Country	Sample	Methods	Key Findings
[19]	United States	driving simulator	driving simulator	Dynamic message signs and arrow pavement markings reduce lane changes.
[20]	China	driving simulator	driving simulator	Toll station guidance signs can improve drivers' lane change behavior.
[13]	Serbia	E70	$M/M/1$, RNNs	RNNs accurately predicted vehicle arrival rates, enhancing model robustness.
[8]	China	Shanghai ring expressway	$M/M/1$, active set	The proposed method reduces queue time by up to 57.4% compared to proportional allocation.
[14]	China	Nanjing toll station	$M/M/1$, enumeration	The benefit decreases initially, then rises with increasing ETC vehicle ratio.
[23]	Egypt	simulation	$M/M/N$, simulation	MTC is inefficient and can cause significant delays to highway traffic.
[24]	Ethiopia	Addis-Adama expressway	$M/M/N$, simulation	Toll service performance depends on the number of servers, service time, and interarrival time.
[25]	United States	New Jerseys expressway	$M/G/1$, fluid mechanics	The vehicle security distance impacts toll booth traffic capacity.
[22]	China	Dongshe toll station	$M/G/N$, PSO-LSTM	The lane scheduling model saves 147,825 RMB annually at Dongshe toll station

However, current research still lacks a comprehensive analysis of how the proportions of different vehicle types and demand flow impact lane choice behavior and has yet to fully consider the interaction between supply and demand. Existing methods typically rely on complex computational models, lacking convenient, efficient lane allocation strategies, which poses challenges for toll station planners and managers. This study thus aims to integrate supply-demand interactions to offer toll station managers a convenient and rapid method for lane allocation.

1.2. Contributions and Paper Outline

The contributions of this study are as follows: First, we propose the heterogeneous traffic flow mixed toll station lane allocation problem (TSLAP for short) and develop a mixed-integer nonlinear bilevel programming problem (MINLBP for short) model to solve it. Second, using the Liulin toll station (Lianhuo Expressway) in Zhengzhou, China as a case study, we validated the model's effectiveness through extensive numerical experiments and analyzed the impact of vehicle composition and arrival rates on system performance, providing valuable management insights. Third, we derived the relaxed analytical solution of the MINLBP model, offering a theoretical explanation for the observed behavior of all ETC-HVs selecting ETC lanes, and providing toll station managers with a quick and efficient lane allocation strategy.

The remainder of this study is organized as follows: Section 2 presents the problem statement, notations, and assumptions. Section 3 details the TSLAP bi-level model and the equivalent single-level model. Section 4 validates the effectiveness of the proposed model through numerical experiments, analyzes the impact of vehicle composition and arrival rate on system performance, and explains the numerical results using the relaxed analytical solution. Section 5 concludes the study and discusses future research directions. Appendix A presents the necessary conditions for the existence of feasible solutions, while Appendix B provides the derivation of the relaxed analytical solution.

2. Problem Statement, Notations, and Assumptions

We consider a bidirectional toll station system (TSS for short) with entrances and exits. Let the set $\boldsymbol{D} = \{1, 2\}$ represent the transport direction of the lanes, indexed by d. The TSS has a total of n_{sum} toll lanes, each of which can use electronic or manual toll collection. The toll collection method is represented by the set $\boldsymbol{I} = \{1, 2\}$, indexed by i. Let n_{di} denote the number of lanes of type i in direction d.

2.1. Study Area

This study focuses on the Liulin Toll Station of the G30 Lianhuo Expressway in China. Located on the outer ring expressway of Zhengzhou, Henan Province, the Liulin Toll Station serves as a critical gateway to the city. Its proximity to major urban arterial roads, the city's northern bus terminal, universities, and densely populated residential areas underscores its significance in meeting substantial population and traffic demands. Figure 2 illustrates the spatial positions of Liulin toll station, urban main roads, and Lianhuo Expressway. As shown in Figure 3, the Liulin Toll Station currently features 14 bidirectional toll lanes. In the entry direction (onto the expressway), there are six lanes, including four inner ETC lanes and two outer MTC lanes. In the exit direction (off the expressway), there are eight lanes, comprising six inner ETC lanes and two outer MTC lanes. The toll lanes at the station are separated by isolation barriers (marked in yellow in Figure 3) and feature extended lane guidance and separation lines (marked in red Figure 3). These measures prevent vehicles entering the toll queue system from freely switching lanes. Figure 2 illustrates the spatial positions of Liulin toll station, urban main roads, and Lianhuo Expressway.

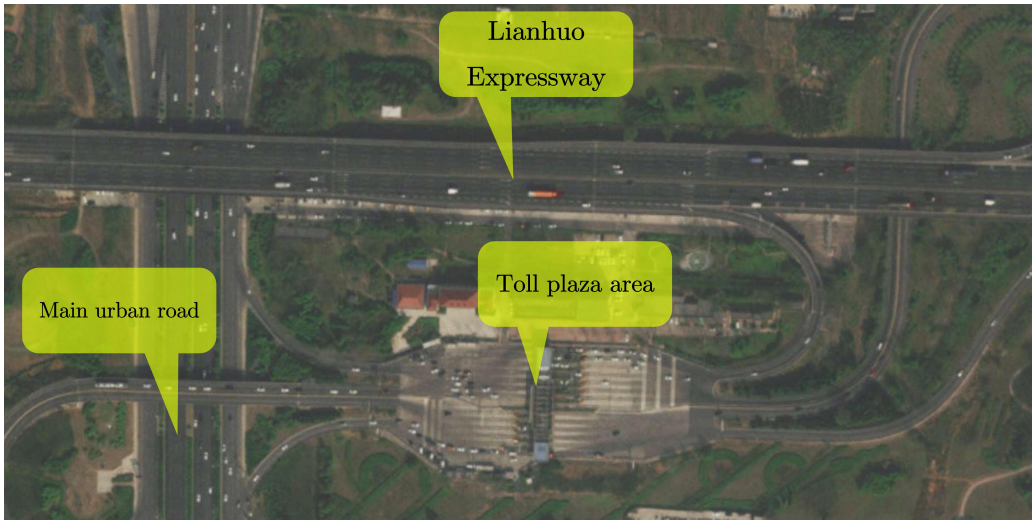

Figure 2. The Liulin toll station and its surrounding roads.

Using the Henan Provincial Expressway Cloud Real-Time Monitoring Platform, video surveillance data from the Liulin Toll Station was collected for 31 October 2024, between 15:30 and 16:30. This time frame was selected for its typical and representative nature: it falls on a Thursday, a midweek working day, and is neither a holiday nor a special workday (e.g., Monday or Friday). The chosen period is also outside peak and off-peak traffic hours.

Vehicles can be classified not only by toll collection method but also by factors such as vehicle size and weight, into passenger cars, buses, and trucks. Clearly, based on this classification, different vehicle types may exhibit varying service times on the same lane type. Following conventional methods, we do not adjust the service rate of the lanes but rather convert vehicle arrival rates (traffic flow) accordingly. Traffic flow data were

standardized using passenger cars as the reference vehicle type, with conversion factors of 1, 2, and 2.5 applied to passenger cars, buses, and trucks, respectively [30]. During this period, the vehicle arrival rates for the entry and exit directions were 23.64 veh/min (1418.5 veh/h) and 21.03 veh/min (1262 veh/h), respectively. The vehicle distribution between the entry and exit directions was similar, with MTC-HVs and ETC-HVs accounting for 0.26 and 0.74 of the traffic. For ETC lanes, the service rates for both entry and exit lanes were similar, at 13.95 veh/min. However, for MTC lanes, the service rates for the entry and exit lanes showed significant differences, at 4.05 veh/min and 2.79 veh/min, respectively. This discrepancy is due to the different processes: at the entry lanes, drivers only need to obtain a highway entry ticket from the toll staff; at the exit lanes, drivers must return the ticket, pay the toll, and sometimes the toll staff needs to provide change. As a result, the service rate for the exit lanes is lower than that for the entry lanes.

Figure 3. Liulin toll station toll plaza.

2.2. Toll Station Queuing System

A toll station is a discrete flow facility. At the entrance lanes of the toll station, vehicles proceed in sequence to either obtain a ticket or record electronically; at the exit lanes, vehicles sequentially make payments to complete the transaction. This creates a queuing process that allows for further analysis based on queueing theory. Due to the physical separation between lanes, vehicles cannot freely switch lanes during queuing, enabling each lane to be considered an independent single-server queuing system. Since this study focuses on solving lane allocation issues at highway toll stations rather than analyzing traffic flow characteristics, traffic volume prediction, or queuing system modeling, and to reduce computational burdens in subsequent model iterations, this study assumes an $M/M/1$ queuing system to describe vehicle arrival and service processes at the toll station.

Let μ_{di} denote the service rate of a single lane of type i in direction d, λ_d denote the arrival rate of vehicles in direction d and λ_{di} denote the vehicle arrival rate for lane type i in direction d. We assume that the service rate of each type of toll lane is independent of the category of vehicles allowed to pass. On the MTC lanes, ETC-HVs and MTC-HVs are treated equivalently; on the ETC lanes, ETC-HVs and CAVs are considered equivalent. Thus, the average passage time Ws_{di} of vehicles in a specific category of lanes in a given direction at a toll station can be expressed as follows:

$$Ws_{di} = \frac{1}{\mu_{di} - \frac{\lambda_{di}}{n_{di}}}, \tag{1}$$

where Ws_{di} includes both service time and queueing time.

Furthermore, the lane allocation scheme at the toll station must ensure the stability of the queuing system, which requires that the service intensity for each type of lane is less than 1. This condition can be mathematically expressed as

$$\frac{\lambda_{di}}{n_{di}\mu_{di}} < 1, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}. \tag{2}$$

It ensures that the system can manage incoming traffic without resulting in congestion, thereby maintaining efficiency.

2.3. Heterogeneous Traffic Flow

Based on the toll collection methods, the toll station faces a mixed traffic flow composed of CAVs, MTC-HVs, and ETC-HVs. Let the set $\boldsymbol{J} = \{1, 2, 3\}$ represent vehicle categories, indexed by j; p_{dj} denotes the proportion of vehicle type j in direction d, satisfying the constraints:

$$\sum_{j \in \boldsymbol{J}} p_j = 1, \tag{3}$$

$$p_j \geqslant 0, \quad \forall j \in \boldsymbol{J}. \tag{4}$$

ETC-HVs have flexibility, allowing them to use either ETC or MTC lanes. Given the short queuing sections at the toll station, ETC-HV users have full visibility of the toll station's traffic conditions prior to lane selection and can leverage their experience to choose the lane that minimizes their passage time. We assume that all ETC-HVs are fully informed about the traffic conditions in the toll station queuing system and are rational in selecting the lane to minimize their passage time. Let θ_{di} represent the proportion of ETC-HVs choosing lane type i in direction d, with the constraints:

$$\sum_{i \in \boldsymbol{I}} \theta_{di} = 1, \quad \forall d \in \boldsymbol{D}, \tag{5}$$

$$\theta_{di} \geqslant 0, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}. \tag{6}$$

In practice, both CAVs and MTC-HVs aim to reduce travel costs by selecting optimal lanes. However, for CAVs on MTC lanes and MTC-HVs on ETC lanes, the travel cost is considered prohibitive. Consequently, CAVs, which rely on automation, are restricted to ETC lanes, while MTC-HVs without ETC equipment must use traditional manually tolled lanes.

Let λ_d denote the vehicle arrival rate in direction d. For a specific type of lane in a given direction, the serviced vehicles include those restricted to that lane type (such as CAVs or MTC-HVs) and a portion of ETC-HVs choosing that lane type. This relationship can be expressed with the following formula:

$$\lambda_{di} = (p_i + p_3 \cdot \theta_{di})\lambda_d, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}. \tag{7}$$

Additionally, to establish a controlled environment in this study, we assume that traffic demand at each toll station in both directions (including vehicle arrival rates and type proportions) is predetermined and fixed, unaffected by changes in lane allocation strategies.

To enhance readability, the notations used in this study are listed in Table 2.

Table 2. List of Notations.

Indices and Sets	
$\boldsymbol{D}$	Set of lane directions
$\boldsymbol{I}$	Set of toll collection methods
$\boldsymbol{J}$	Set of vehicle categories
d	Direction: 1 for entrance, 2 for exit
i	Lane type: 1 for ETC lane, 2 for MTC lane
j	Vehicle type: 1 for CAVs, 2 for MTC-HVs, and 3 for ETC-HVs
c	Indicates that the decision variable is relaxed
$*$	Indicates that variables are treated as parameters
Parameters	
μ_{di}	The service rate of a single lane of type i in direction d
n_{sum}	The total number of lanes at the highway toll station
p_{dj}	The proportion of vehicle type j in direction d
θ_{di}	The proportion of ETC-HVs choosing the lanes of type i in direction d
λ_d	The arrival rate of vehicles in direction d
Ws_{di}	The average passage time of vehicles in lane type i in direction d
k	Lagrange multiplier
Decision variables	
n_{di}	The number of lanes of type i in direction d
λ_{di}	The vehicle arrival rate for lane type i in direction d

3. Methodology

In this section, we first formulate the bilevel programming model for TSLAP, capturing the interplay between supply and demand at toll stations. The upper-level model aims to minimize total vehicle travel time by optimizing the allocation of different types of lanes in each direction at the toll station under a given traffic demand. The lower-level model adopts the user equilibrium principle to examine the lane selection behavior of ETC-HVs under a specified lane allocation scheme. Second, by applying the Lagrangian multiplier method, we incorporate the Karush–Kuhn–Tucker (KKT for short) conditions of the lower-level model into the upper level, transforming the bilevel programming model of TSLAP into an equivalent single-level programming model.

3.1. Optimization Model

Assuming that non-highway routes between toll stations are disregarded, we can treat each toll station within the highway network as an independent entity. This approach allows for lane allocation optimization to be conducted separately at each toll station. To enhance readability, we illustrate the relationships within the bi-level programming model for lane allocation at highway toll stations, as shown in Figure 4.

3.1.1. Highway Toll Station Lane Allocation

In accordance with Section 2, the lane allocation problem at toll stations under heterogeneous traffic flows is presented from the perspective of traffic planners and managers. They adhere to the principle of system optimization by determining and adjusting the number of ETC lanes in the queueing system to minimize total travel time. We use the subscripts "up" and "low" to differentiate between the upper and lower models in the bilevel programming framework. Let Z_{up} represent the objective function; thus, the objective function for the toll station lane allocation problem can be expressed as follows:

$$\min_{n_{di}} Z_{up} = \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \lambda_{di} Ws_{di} \tag{8}$$

subject to

$$\sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} n_{di} = n_{sum}, \tag{9}$$

$$n_{di} > \frac{\lambda_{di}}{\mu_{di}}, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}, \lambda_{di} \neq 0, \tag{10}$$

$$n_{di} \in \mathbb{N}, \tag{11}$$

where Equation (9) corresponds to the total number of lanes constraint; Equation (10) is derived from Equation (2) and ensures the stable operation of the toll station, confirming that Ws_{di} is a non-negative value; Equation (11) ensures each lane number is an integer.

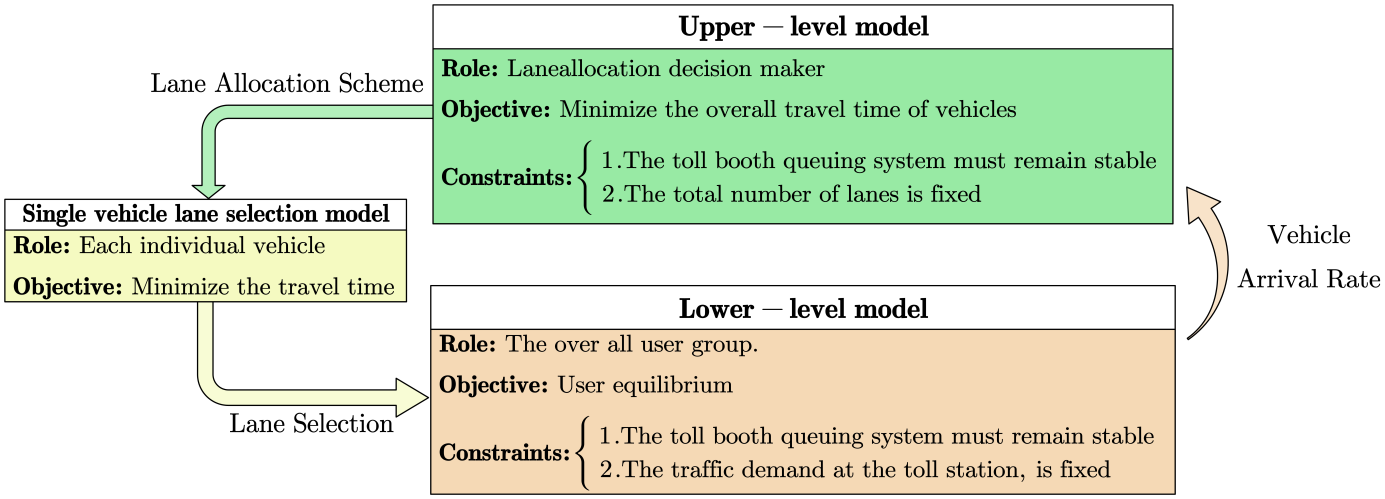

Figure 4. Bi-level programming model for lane allocation at highway toll stations.

3.1.2. Multi-Class User Equilibrium

Due to the traffic conditions inherent in the toll plaza queue system, ETC-HVs do not experience temporal or spatial invisibility. Consequently, they are already aware of the traffic situation at the toll plaza before selecting a lane. Moreover, based on their experience, ETC-HVs can identify the lane that minimizes their time spent at the toll plaza. Therefore, we assume that all users are rational and have a clear understanding of the traffic conditions within the toll plaza queue system. Under a given toll lane allocation strategy, we determine the lane choice of ETC-HVs according to the user equilibrium principle.

It is important to note that when the lane allocation strategy is fixed, vehicles entering and exiting the toll plaza do not compete with each other, as users can only access lanes corresponding to their direction of travel and compete solely with vehicles moving in the same direction. Furthermore, even when considering dynamic demand, vehicles during different time intervals do not compete with one another. Thus, we can treat the lane choices of users in various directions and time intervals as independent and analyze them separately.

If we regard the highway toll plaza queue system as an equivalent traffic network, then for an ETC-HV user traveling in one direction, there are only two available paths: the ETC lane and the MTC lane. In this scenario, we need not adhere strictly to the conventional objective function forms of traffic network user equilibrium models.

For an individual ETC-HV, the lane selection is guided by the principle of minimizing travel time. Specifically, if the travel time on the MTC lane is less than that on the ETC lane, the vehicle will choose the MTC lane; otherwise, it will opt for the ETC lane. This relationship can be expressed mathematically as follows:

$$choice = \begin{cases} ETC, & \text{if } Ws_{d1} \leqslant Ws_{d2}, \\ MTC, & \text{if } Ws_{d1} > Ws_{d2}. \end{cases} \tag{12}$$

According to Equation (1), the average travel time Ws_i for lane type i increases as the user arrival rate λ_i for that lane type rises. Consequently, when an individual ETC-HV opts for a lane with a shorter travel time, it results in an increase in the average travel time for that lane type. Therefore, the lane selection behavior of ETC-HVs tends to equalize the average travel times of both the MTC and ETC lanes. This relationship can be represented mathematically as follows:

$$\min_{\lambda_{di}} |Ws_{d1} - Ws_{d2}| \tag{13}$$

subject to

$$\lambda_{di} < \mu_{di} n_{di}, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}, n_{di} \neq 0, \tag{14}$$

$$\lambda_{di} \geqslant p_i \lambda_d, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}, \tag{15}$$

$$\sum_{i \in \boldsymbol{I}} \lambda_{di} = \lambda_d, \quad \forall d \in \boldsymbol{D}, \tag{16}$$

where Equation (14) is derived from Equation (2); Equations (15) and (16) are derived from Equations (3)–(6).

However, the objective function in Equation (13) contains absolute values, and the expression for Ws_{di} is a fractional structure. Given that $Ws_{di} > 0$, for computational and analytical convenience, we can reframe the model using $\frac{n_{d1} n_{d2}}{Ws_{di}}$. This indicates that an individual ETC-HV user will select the lane type that minimizes $\frac{n_{d1} n_{d2}}{Ws_{di}}$. Thus, the selection behavior of a single ETC vehicle can be expressed as follows:

$$choice = \begin{cases} ETC, & \text{if } \frac{n_{d1} n_{d2}}{Ws_{d2}} \leqslant \frac{n_{d1} n_{d2}}{Ws_{d1}}, \\ MTC, & \text{if } \frac{n_{d1} n_{d2}}{Ws_{d2}} > \frac{n_{d1} n_{d2}}{Ws_{d1}}. \end{cases} \tag{17}$$

Let Z_{low} denote the objective function for the lower-level planning. The objective function for the multi-user equilibrium problem at the toll plaza can be expressed as follows:

$$\min_{\lambda_{di}} Z_{low} = \left(\frac{n_{d1} n_{d2}}{Ws_{d1}} - \frac{n_{d1} n_{d2}}{Ws_{d2}} \right)^2. \tag{18}$$

By substituting Equation (1) into Equation (18) and rearranging, we obtain the new objective function as follows:

$$\min_{\lambda_{di}} Z_{low} = \left(n_{d1} n_{d2} (\mu_{d1} - \mu_{d2}) - n_{d2} \lambda_{d1} + n_{d1} \lambda_{d2} \right)^2 \tag{19}$$

subject to the following constraints: Equations (14)–(16).

Among the aforementioned constraints, Equation (14) ensures the stable operation of the toll plaza, while guaranteeing that Ws_{di} remains as a positive value. Equation (15) establishes the lower limit for the vehicle arrival rates of each lane type, specifically the case where $\theta_{di} = 0$, indicating that vehicles in the ETC lane consist solely of CAVs, or that vehicles in the MTC lane consist exclusively of MTC-HVs. Finally, Equation (16) represents the constraint on total traffic demand.

3.2. Single-Level Programming Model for the TSLAP

Since both the upper and lower-level problems are convex, the KKT conditions of the lower-level problem can be used to transform the bi-level problem into a single-level problem.

Constraint (14) of the lower-level model is equivalent to constraint (10) of the upper-level model. Thus, when constructing the Lagrangian function for the lower-level model, constraint (14) does not need to be included as a complementary slackness condition in the objective function, as it would introduce redundant constraints into the single-level formulation of TSLAP.

The Lagrangian function of the lower-level problem $\min_{\lambda_{di}} L_{low}$ is constructed as follows:

$$\min_{\lambda_{di}}(n_{d1}n_{d2}(\mu_{d1}-\mu_{d2})-n_{d2}\lambda_{d1}+n_{d1}\lambda_{d2})^2+\sum_{i\in \boldsymbol{I}}k_{di}(p_i\lambda_d-\lambda_{di})+k_{d3}\left(\sum_{i\in \boldsymbol{I}}\lambda_{di}-\lambda_d\right),$$

where k_{di} and k_{d3} represents the Lagrange multiplier.

The KKT conditions of the lower-level model are as follows:

$$\frac{\partial L_{low}}{\partial \lambda_{di}}=0, \quad \forall d\in \boldsymbol{D}, i\in \boldsymbol{I}, \tag{20}$$

$$k_{di}(p_i\lambda_d-\lambda_{di})=0, \quad \forall d\in \boldsymbol{D}, i\in \boldsymbol{I}, \tag{21}$$

$$k_{d3}\left(\sum_{i\in \boldsymbol{I}}\lambda_{di}-\lambda_d\right)=0, \quad \forall d\in \boldsymbol{D}, \tag{22}$$

$$k_{di}\geqslant 0, \quad \forall d\in \boldsymbol{D}, i\in\{1,2,3\}, \tag{23}$$

along with Equations (14)–(16).

By incorporating Equations (14)–(16) and (20)–(23) into the constraints of the upper-level model, the single-level programming formulation of TSLAP is constructed.

4. Numerical Experiments

In this section, we use the data from Section 2.1 as parameters for the model in Section 3.2 and conduct numerical experiments using the commercial solver Gurobi.

4.1. Computational Performance

A comparison between the current state and the optimized results is shown in Figure 5 and Table 3.

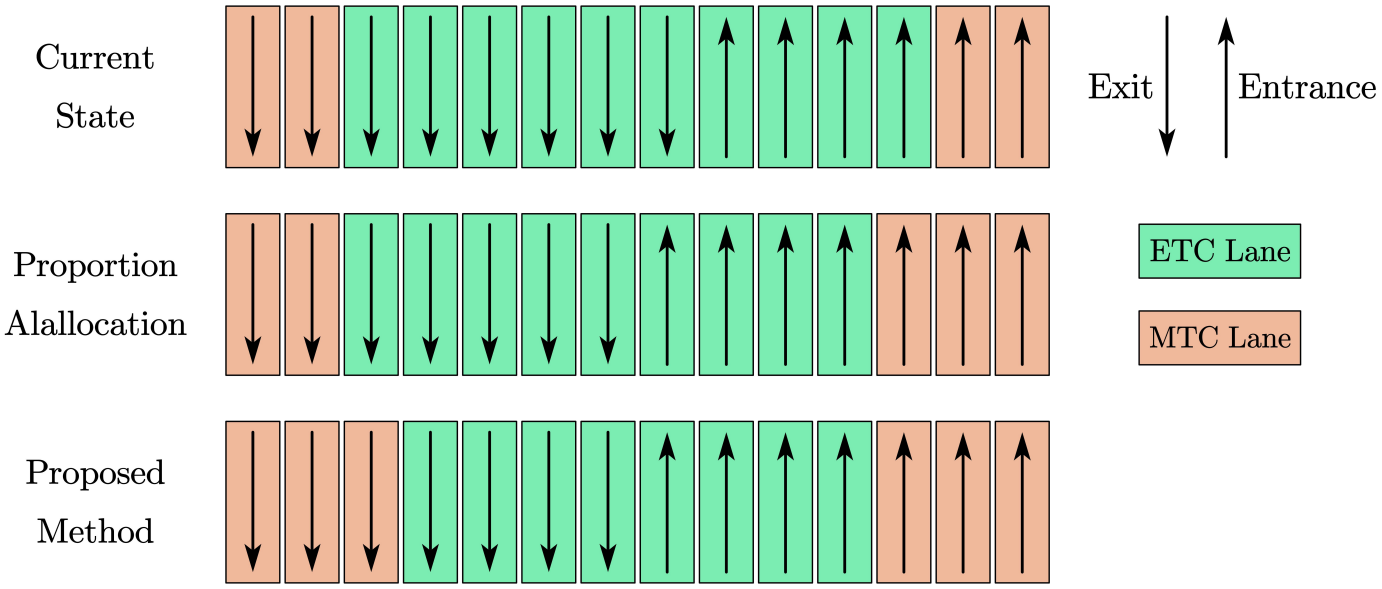

Figure 5. Lane allocation results under the current situation, proportional allocation, and proposed method.

Table 3. Comparison of the current situation, proportional allocation, and the proposed method.

Lane Allocation and System Performance	Current State		Proportional Allocation		Proposed Method	
	Entry	Exit	Entry	Exit	Entry	Exit
Number of ETC lanes (lanes)	4	6	5	5	3	3
Number of MTC lanes (lanes)	2	2	2	2	4	4
Traffic intensity of ETC lanes	0.31	0.19	0.25	0.22	0.42	0.37
Traffic intensity of MTC lanes	0.76	0.98	0.76	0.98	0.38	0.49
Total travel time (minutes)	106.95		106.87		10.22	

4.2. The Impact of Vehicle Composition and Arrival Rate

In this section, we will examine the impact of two factors: vehicle arrival rate and vehicle composition. We assume a constant service rate and identical vehicle distribution at the entry and exit lanes. The proportions of different vehicle types p_{dj} and the vehicle arrival rate λ_d will be adjusted to explore their impact.

Given that vehicles are divided into three types, with their proportions summing to one, determining the proportions of any two types allows for the derivation of the third. For analytical clarity, rather than simultaneously varying two vehicle proportions, we hold one type's proportion constant while adjusting the other to investigate the impact of vehicle type distribution. Figure 6 presents a series of curves representing the variation in the optimal total toll station travel time as the proportion of ETC-HVs changes, while either CAV or MTC-HV proportions remain fixed. The difference in CAVs or MTC-HVs proportion between adjacent curves is 0.05. To enhance readability, the curve colors darken progressively as the proportion of CAVs or MTC-HVs increases. Figure A1 shows the optimal solution corresponding to Figure 6.

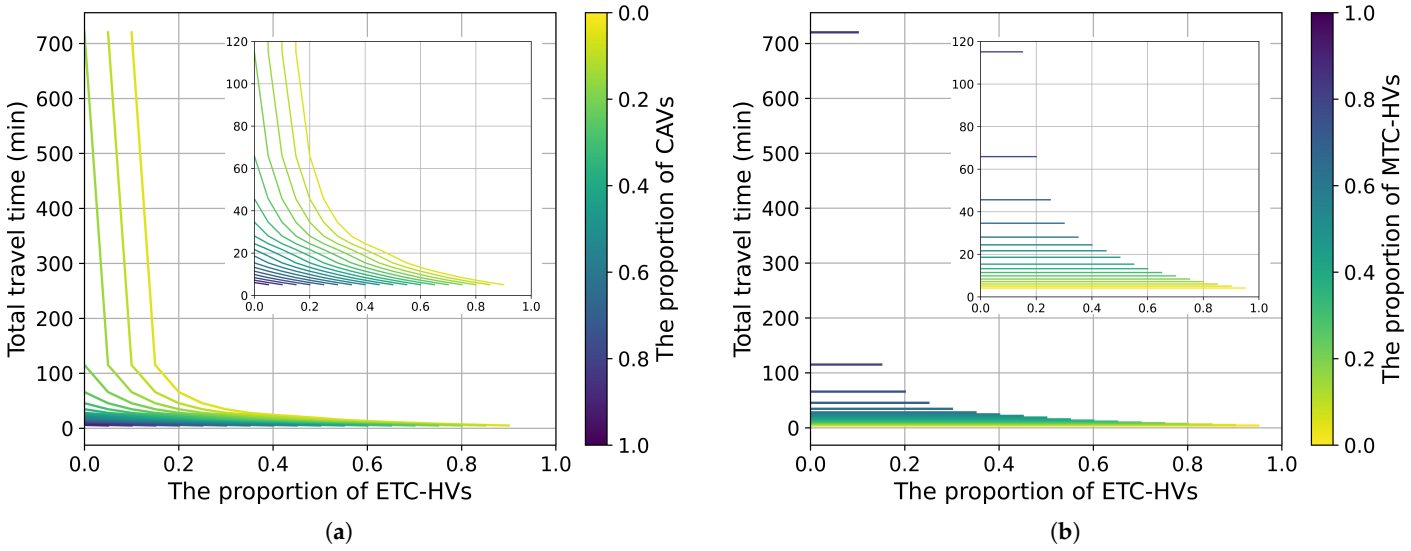

Figure 6. Total travel time as ETC-HVs proportion varies with fixed CAVs or ETC-HVs proportion. (**a**) The CAVs proportion is fixed first; (**b**) the ETC-HVs proportion is fixed first.

To analyze the impact of vehicle proportions more effectively, we first fix the vehicle arrival rate in the exit direction and then adjust the arrival rate in the entry direction, which varies from 0 to 50. Figure 7 presents a series of curves, forming a set where each curve illustrates the variation in total toll station travel time as the entry-direction vehicle arrival rate changes, with a fixed exit-direction rate. The difference in exit-direction arrival rate between adjacent curves is 5 veh/min. For clarity, curve colors transition from light to

dark as the exit-direction arrival rate increases. Figure A2 shows the optimal solution corresponding to Figure 7.

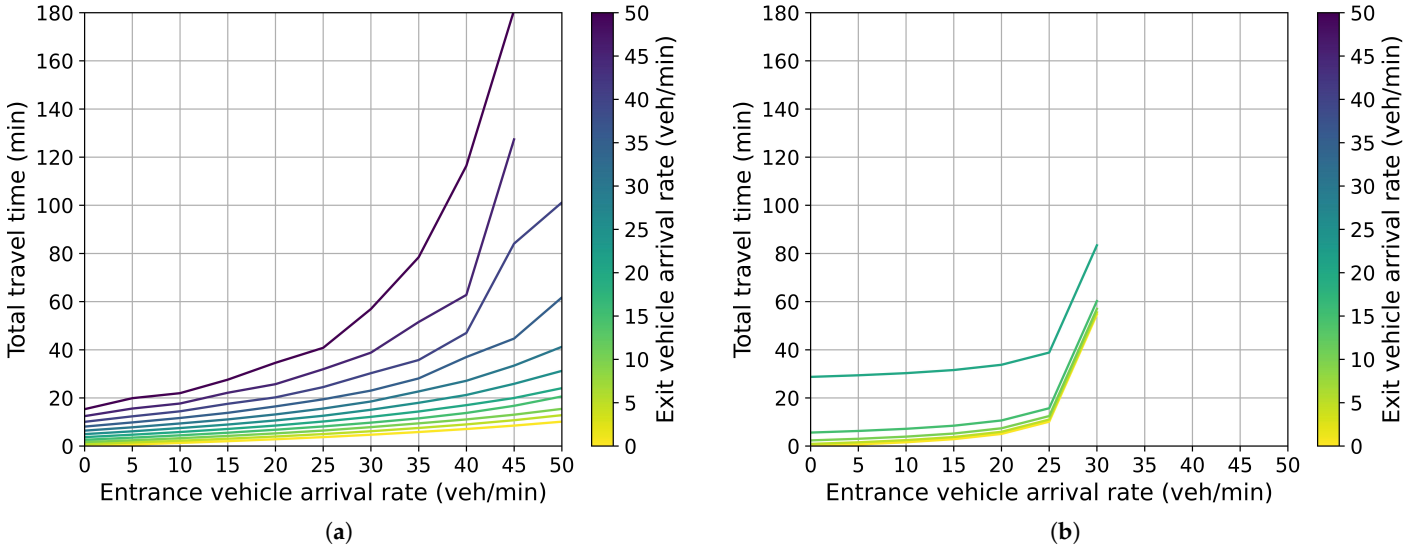

Figure 7. Total travel time variation with entry-direction flow under fixed exit-direction flow. (**a**) Lane allocation under optimization model; (**b**) current lane allocation.

4.3. Discussion

In this section, we discuss the numerical experiment results and analyze the effectiveness of the proposed model. We also identify which proportion among CAVs, MTC-HVs, and ETC-HVs is the most critical. Additionally, we explore the potential of the lane allocation adjustment strategy.

4.3.1. The Proposed Model Enhances Toll Station Performance with Computational Efficiency

Based on Table 3, it is evident that the current allocation of ETC lanes is excessive, leading to the concentration of MTC-HV vehicles on a limited number of MTC lanes. The total system throughput time at the toll station is influenced not only by the service rate of individual lanes but also by the overall lane distribution and resource scheduling. The overloading of MTC lanes results in decreased system efficiency.

After optimizing the lane allocation scheme using our model, the total travel time of the toll station was reduced by 90.44% compared to the current situation and proportional allocation method. This demonstrates that our model is effective in significantly improving toll station performance under heterogeneous traffic flow conditions, maximizing the potential of ETC lanes to reduce congestion.

These findings align with prior studies that have highlighted the critical role of lane allocation in minimizing travel time at toll stations [8,12,22]. Additionally, our model's approach differs from traditional models by incorporating bi-level programming and Lagrangian relaxation, allowing for a more accurate representation of heterogeneous traffic flow. Compared to simpler proportional allocation methods, our allocation strategy achieves greater reductions in system throughput time. Moreover, by deriving the relaxed analytical solution, the computational convenience of this allocation strategy is significantly enhanced.

4.3.2. MTC-HVs Proportion: The Most Critical Factor Among Vehicle Types

In Figure 6b, each curve is horizontal, indicating that when the proportion of MTC-HVs is fixed, changes in the proportion of ETC-HVs (with the CAVs proportion simultaneously varying) have no effect on total passage time. Thus, since all MTC-HVs choose the ETC

lane, CAVs and MTC-HVs can be considered homogeneous in terms of total passage time or optimal solutions.

In Figure 6a, the curve with a low CAVs proportion (lighter color) does not intersect the vertical axis, meaning that at this CAVs proportion, the ETC-HVs proportion cannot be zero. A low ETC-HVs proportion combined with a high MTC-HVs proportion will cause system failure. In Figure 6b, the rightmost point of each curve represents the maximum value of ETC-HVs, i.e., $1 - p_2$, indicating that no line exists where p_2 is less than 0.1. These findings indicate that there is a threshold for the MTC-HVs proportion. The threshold implies that, given the current traffic demand (vehicle arrival rate), the MTC-HVs proportion should be reduced to below this threshold; otherwise, the toll station queuing system will collapse, leading to severe congestion.

In Figure 6a, the rightmost point on each curve represents the maximum possible proportion of ETC-HVs, i.e., $1 - p_1$. The minimum value occurs when the ETC-HVs proportion is at its maximum, where the sum of CAVs and ETC-HVs proportions equals 1, and no MTC-HVs exist. In Figure 6b, the curve with zero MTC-HVs proportion (brightest color) shows the shortest total passage time. This indicates that the queuing system has the lowest total passage time when no MTC-HVs are present.

Figure A1 illustrates the optimal solution corresponding to the optimal value curve in Figure 6. In Figure 6b, we observed that when the proportion of MTC-HVs is fixed, the optimal value remains unchanged, leading us to suspect that the optimal value is constant. Evidently, this hypothesis is confirmed in Figure A1. This result differs somewhat from intuitive expectations. To explain this intriguing numerical finding, we derive the relaxed analytical solution of the bi-level programming model using the Lagrangian relaxation method in Appendix B, as shown in Equation (A5).

In conclusion, the MTC-HVs proportion should be the most critical factor in toll station lane planning and management.

4.3.3. Dynamically Adjusting Lane Allocation Exhibits Significant Potential

Under the same vehicle arrival rates at both the entry and exit points, adjusting the toll booth lane allocation based on our optimization model consistently outperforms the current system, with increasingly pronounced improvements as traffic volume rises. Moreover, the number of curves in Figure 7a exceeds that in Figure 7b, and the maximum value on the x-axis of Figure 7a is greater than that of Figure 7b. These findings indicate the effectiveness of our model, showing that lane allocation adjustments based on our optimization approach enable toll booths to better manage high traffic volumes. Additionally, dynamically adjusting lane allocation in response to changes in vehicle arrival rates proves to be a more effective supply strategy than maintaining a static allocation.

Certainly, we must recognize that dynamic lane allocation adjustment faces multiple challenges in practical implementation. First, it relies heavily on the acquisition and processing of real-time data, requiring toll stations to possess efficient data collection and analysis capabilities. However, under high traffic volumes or complex traffic conditions, data delays and technological limitations may affect the timeliness of adjustments. Secondly, accurately predicting changes in traffic flow is inherently difficult, as traffic volumes are influenced by numerous factors such as weather, accidents, and holidays, thereby increasing the complexity of making timely adjustments. Furthermore, existing tolling infrastructure and hardware (e.g., tidal lanes and variable toll rate signs) may not be capable of quickly or flexibly adapting, as these adjustments often involve high costs and operational difficulties. Operators also need to make rapid decisions, which necessitates specialized training and efficient coordination, particularly during peak periods, where slow response times could become a bottleneck. Another challenge is the compatibility issue between different sys-

tems—such as traffic management, tolling, and monitoring systems—whose technological disparities could complicate implementation. While dynamic adjustments can improve efficiency, they also require significant financial investment in facility upgrades, technological inputs, and ongoing maintenance, which could be a substantial burden for toll stations with limited resources. Additionally, regional differences play a significant role, as toll stations in various areas may differ in traffic flow, infrastructure, and available resources. In regions with low traffic or minimal demand, static lane allocation may already suffice, making dynamic adjustment unnecessary or infeasible. In conclusion, while dynamic lane allocation adjustment has the potential to significantly enhance efficiency in highway toll stations, its complexity, costs, and implementation challenges make it difficult to achieve in many cases. However, with the advancement of smart tolling systems, we believe the application potential of this approach is immense.

5. Conclusions and Future Research

This study investigates the TSLAP under heterogeneous traffic flow, considering the lane selection rights of different vehicle types. Based on the user equilibrium principle and queueing theory, the lane selection behavior of ETC-HVs is modeled. The TSLAP is formulated as a mixed-integer nonlinear bilevel optimization problem, aiming to minimize the total passage time of the toll station by optimizing the number of ETC lanes at the station's entrances and exits. Since both the upper and lower-level problems are convex, the KKT conditions of the lower-level problem are used to transform the bilevel problem into an equivalent single-level optimization. The problem is solved numerically using the commercial solver Gurobi, and a relaxed analytical solution is derived using a parameterized iterative method. The results show that the proposed approach significantly improves the performance of toll station queuing systems under heterogeneous traffic flow, reducing the total passage time by 90.44% compared to the current system. Impact analysis and comparative experiments reveal that the proportion of MTC-HVs has the greatest impact on the optimal total passage time while increasing the market penetration of CAVs or ETC-HVs helps manage high traffic demand. Additionally, dynamically adjusting lane allocation in response to changes in vehicle arrival rates proves to be a more effective strategy than maintaining a static allocation. The developed model and obtained results can serve as a basis for planning and managing toll station lane allocation in heterogeneous traffic flow environments.

However, this study has several limitations. First, the study assumes that the toll station queuing system follows an M/M/1 model. On one hand, each lane is treated as an independent single-server queuing system, which is a reasonable assumption for the Liulin toll station, as its lanes are equipped with barriers and extended vehicle separation guidelines that prevent vehicles from switching lanes freely during the queuing process. However, other toll stations may allow vehicles to switch lanes. For toll stations that allow lane switching, a multi-server queueing system could potentially be used. However, considering the practical situation, it is quite difficult for vehicles to change lanes continuously before the toll plaza—such as from the outermost lane to the innermost lane—and during congestion, the spacing between vehicles does not allow lane changes. This may not align with the characteristics of a multi-server queueing system, which typically allows free switching between servers. Additionally, lane switching would disrupt the toll plaza system, potentially affecting the speed of vehicles behind. We suggest that for toll stations where lane switching is allowed, simulation techniques may be a better approach for modeling such systems. On the other hand, this study focuses on the lane allocation problem at toll stations and assumes a simple M/M/1 queuing system, characterized by Poisson arrivals and exponential service times. In practice, the situation is more complex

and may change over time, requiring more sophisticated models. Second, while numerical experiments in this study demonstrate that dynamic lane allocation can effectively manage traffic fluctuations, it is important to recognize that real-world applications involve more complex operational processes. With the increasing sophistication of traffic management systems, we believe dynamic adjustment holds significant application potential. Third, this study assumes that ETC-HV users are fully rational in their lane selection, always choosing the lane with the shortest waiting time. However, real-world behavior is likely more complex, with factors such as driver preferences, experience, and external influences affecting lane choice.

Future research will focus on developing more realistic toll station service models, with particular attention given to the impacts of toll station geometry, user behavior, and vehicle-following effects on service efficiency. The geometric design of toll stations not only affects vehicle mobility but is also closely linked to service rate. Vehicle headway is typically influenced by vehicle type characteristics and the following behavior of preceding vehicles. Consequently, service rate models need refinement to account for the combined effects of diverse vehicle types, tolling methods, driving behaviors, and geometric layouts. By developing more refined service models, more precise lane allocation can be achieved.

Author Contributions: Conceptualization, Z.C., T.R. and M.X.; methodology, Z.C. and T.R.; software, Z.C. and T.R.; validation, Z.C., T.R. and M.X.; formal analysis, Z.C. and T.R.; investigation, Z.C. and T.R.; resources, Z.C. and T.R.; data curation, Z.C. and T.R.; writing—original draft preparation, Z.C., T.R. and M.X.; writing—review and editing, Z.C., T.R. and M.X.; visualization, Z.C. and T.R.; supervision, M.X.; project administration, M.X.; funding acquisition, M.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. PolyU 15221821).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

TSLAP	Toll station lane allocation problem
MINLBP	Mixed-integer nonlinear bilevel programming
ETC	Electronic toll collection
MTC	Manual toll collection
CAVs	Connected Autonomous Vehicles
TTS	Toll station system
KKT	Karush–Kuhn–Tucker

Appendix A. Necessary Condition for the Existence of Feasible Solutions

According to Equations (10), (11) and (14), a feasible solution must satisfy the following condition:

$$n_{di} \geq \lceil \frac{\lambda_{di}}{\mu_{di}} \rceil, \quad \forall d \in \boldsymbol{D}, i \in \boldsymbol{I}, \tag{A1}$$

where the symbol $\lceil \rceil$ denotes the ceiling function, which rounds a number up to the nearest integer greater than or equal to it.

Since $\lambda_{di} = 0$ cannot hold for all i simultaneously, and Equation (9), the necessary condition for the existence of a feasible solution is

$$n_{\text{sum}} \geq \min\left(\sum_{d\in \boldsymbol{D}}\sum_{i\in \boldsymbol{I}} \lceil \frac{\lambda_{di}}{\mu_{di}} \rceil\right). \tag{A2}$$

Since $\mu_{d1} > \mu_{d2}$, the following expression

$$\sum_{d\in \boldsymbol{D}}\sum_{i\in \boldsymbol{I}} \lceil \frac{\lambda_{di}}{\mu_{di}} \rceil \tag{A3}$$

decreases as the proportion of ETC-HVs choosing the ETC lanes increases. Therefore, when all ETC-HVs choose the ETC lanes, i.e., $\theta_{d1} = 1, \theta_{d2} = 0$, expression (A3) reaches its minimum value.

Thus, the necessary condition for the existence of a feasible solution is given by the following expression:

$$n_{\text{sum}} \geq \sum_{d\in \boldsymbol{D}}\left(\lceil \frac{p_1 + p_3}{\mu_{d1}}\lambda_d \rceil + \lceil \frac{p_2}{\mu_{d2}}\lambda_d \rceil\right). \tag{A4}$$

It should be noted that the parameters constituting this necessary condition are all given and are not affected by the lane allocation scheme or the choices made by ETC-HV users. Therefore, before solving the model, we can make a preliminary judgment on the existence of a feasible solution using Equation (A4). Since the problem becomes meaningless when a feasible solution does not exist, in this paper, we will only consider the case where a feasible solution exists, treating Equation (A4) as a given basic condition.

Appendix B. The Relaxed Solution of the TSLAP

In this section, we first obtain the relaxed analytical solution to the upper-level problem using the Lagrange multiplier method and verify its feasibility by applying the necessary conditions for feasible solutions presented in Appendix A. Subsequently, we derive the relaxed solution to the bilevel programming problem by combining the relaxed solution of the upper-level problem with a parameter iteration method. In fact, the relaxed solution of the upper-level problem is equivalent to that of the bilevel programming problem. The following presents the detailed theoretical derivation process.

Appendix B.1. Relaxed Solution of the Upper-Level Problem

This section derives the relaxed solution for the upper-level problem, specifically proving Proposition A1. This serves as the foundation for deriving the relaxed solution of the bilevel programming problem.

Proposition A1. *The relaxed analytical solution to the upper-level problem is*

$$n_{di}^{c} = \frac{\frac{\lambda_{di}}{\mu_{di}}}{\sum_{d\in \boldsymbol{D}}\sum_{i\in \boldsymbol{I}} \frac{\lambda_{di}}{\mu_{di}}} n_{\text{sum}}, \tag{A5}$$

where c denotes continuous.

Proof. To make the problem easier to solve, we first relax the integer constraint on the decision variable n_{di} in the upper-level model in order to attempt to find its relaxed solution. The objective function (8) can be rearranged into the following form:

$$\begin{aligned} \min_{n_{di}} Z_{up} &= \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \lambda_{di} W s_{di} \\ &= \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\lambda_{di}}{\mu_{di} - \frac{\lambda_{di}}{n_{di}}} \\ &= \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\lambda_{di}}{\mu_{di}} + \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\left(\frac{\lambda_{di}}{\mu_{di}}\right)^2}{n_{di} - \frac{\lambda_{di}}{\mu_{di}}}. \end{aligned} \tag{A6}$$

We use Z_{up}^c to represent the objective function of the relaxed problem after removing the integer constraint on the upper-level decision variables. Next, we will use the Lagrange multiplier method to solve the relaxed problem and verify that the solution lies within the feasible region.

The relaxed problem of the upper-level optimization is equivalent to

$$L_{up}^c = \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\lambda_{di}}{\mu_{di}} + \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\left(\frac{\lambda_{di}}{\mu_{di}}\right)^2}{n_{di} - \frac{\lambda_{di}}{\mu_{di}}} + k\left(\sum_{d\in \boldsymbol{D}}^{2} \sum_{i\in \boldsymbol{I}} n_{di}^c - n_{\text{sum}}\right). \tag{A7}$$

subject to Equation (10), where k is the Lagrange multiplier. Note that we have only introduced equality constraints into the objective function and have not included inequality constraints.

In the upper-level problem, the decision variable λ_{di} of the lower-level problem can be considered as a parameter. We use the asterisk notation (*) to distinguish between the parameter and variable forms.

From $\frac{\partial L_{up}^c}{\partial n_{di}} = 0$, we obtain the candidate optimal solution of Equation (A7), which is Equation (A5) (the term "candidate" means that this solution has not yet undergone a feasibility check).

Check whether Equation (A5) satisfies the inequality constraint (10):

$$n_{di}^c - \frac{{\lambda_{di}}^*}{\mu_{di}} = \frac{\frac{{\lambda_{di}}^*}{\mu_{di}}}{\sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{\lambda_{di}}{\mu_{di}}} \left(n_{\text{sum}} - \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}}\right). \tag{A8}$$

According to the constraint of the lower-level optimization (Equation (14)), it is known that

$$\sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} < \sum_{d\in \boldsymbol{D}} \sum_{i\in \boldsymbol{I}} {n_{di}^c}^* = n_{\text{sum}}. \tag{A9}$$

By combining Equations (A9) and (A10), it can be proven that Equation (A5) satisfies constraint (10). Therefore, Proposition A1 is proven. □

Appendix B.2. Solving Relaxed Bilevel Programming Problems Using Iterative Methods

When solving a bilevel programming problem, iterative methods are generally an effective approach for tackling such complex issues. Bilevel programming involves two levels of optimization problems, where the upper-level problem (leader) typically influences the decision-making of the lower-level problem (follower). Since the lower-level problem usually depends on the decision variables of the upper-level problem, the core challenge in solving a bilevel programming problem lies in how to effectively interact between these

two levels. The following are the steps to solve the two-layer programming problem based on the iterative method:

Step 1: Initialization

Set the initial upper-level decision variable (e.g., ${n_{di}^c}^{(0)}$), and solve the lower-level problem to find the optimal solution (e.g., ${\lambda_{di}}^{(0)}$) based on these initial decision variables.

Step 2 to 4: Iterative Process

Step 2: In the m-th iteration, solve the lower-level problem using the current upper-level decision variable ${n_{di}^c}^{(m)}$, and obtain the solution ${\lambda_{di}}^{(m)}$.

Step 3: Based on the solution ${\lambda_{di}}^{(m)}$ of the lower-level problem, use the result to update the upper-level decision variable ${n_{di}^c}^{(m+1)}$.

Step 4: Check for convergence. If the convergence criteria are met (e.g., the change in the solution is sufficiently small), stop the iteration; otherwise, return to Step 2 and continue the iteration.

We may initialize the upper-level problem with the solution ${n_{di}^c}^{(0)}$ given by Equation (A5). Since λ_{di} serves as both a parameter in the upper-level model and a decision variable in the lower-level model, Equation (A5) cannot be directly applied to the lower level to avoid altering the original lower-level optimization problem. Assume that all vehicles choose the ETC lane, i.e., $\theta_{d1} = 1$, $\theta_{d2} = 0$. In this case, Equation (A5) represents the following set of initial feasible solutions:

$$ {n_{di}^c}^{(0)} = \frac{\frac{{\lambda_{di}}^*}{\mu_{di}} n_{\text{sum}}}{\sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}}} \tag{A10} $$

subject to

$$ \begin{cases} {\lambda_{d1}}^* = (p_1 + p_3)\lambda_d, \\ {\lambda_{d2}}^* = p_2 \lambda_d. \end{cases} \tag{A11} $$

Substituting the initial solution ${n_{di}^c}^{(0)}$ into the lower-level objective function (Equation (18)), we do not immediately solve it. Instead, we first prove Proposition A2, thereby avoiding the need for a case analysis of its monotonicity.

Proposition A2. *With the initial lane allocation ${n_{di}^c}^{(0)}$, the average travel time on ETC lanes is always less than on MTC lanes, as expressed by the following:*

$$ \frac{1}{{Ws_{d1}}^{(0)}} - \frac{1}{{Ws_{d2}}^{(0)}} > 0. \tag{A12} $$

Proof. By introducing Equation (A10), we obtain

$$ \frac{1}{{Ws_{d1}}^{(0)}} - \frac{1}{{Ws_{d2}}^{(0)}} = \mu_{d1} - \mu_{d2} - \frac{1}{n_{\text{sum}}} \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} \left(\frac{\mu_{d1}\lambda_{d1}}{{\lambda_{d1}}^*} - \frac{\mu_{d2}\lambda_{d2}}{{\lambda_{d2}}^*} \right), \tag{A13} $$

where the asterisk (*) denotes a parameter, rather than a decision variable in the lower-level problem.

Therefore, proving Equation (A12) is equivalent to proving

$$ \mu_{d1} - \mu_{d2} > \frac{1}{n_{\text{sum}}} \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} \left(\frac{\mu_{d1}\lambda_{d1}}{{\lambda_{d1}}^*} - \frac{\mu_{d2}\lambda_{d2}}{{\lambda_{d2}}^*} \right). \tag{A14} $$

Considering Equation (16), it follows that $\frac{\mu_{d1}\lambda_{d1}}{{\lambda_{d1}}^*} - \frac{\mu_{d2}\lambda_{d2}}{{\lambda_{d2}}^*}$ increases as λ_{d1} increases.

Since $\lambda_{d1} \leqslant {\lambda_{d1}}^*$, it follows that

$$\frac{1}{n_{\text{sum}}} \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} \left(\frac{\mu_{d1}\lambda_{d1}}{{\lambda_{d1}}^*} - \frac{\mu_{d2}\lambda_{d2}}{{\lambda_{d2}}^*} \right) \leqslant \frac{1}{n_{\text{sum}}} \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} (\mu_{d1} - \mu_{d2}). \tag{A15}$$

Based on Equation (A9) and the condition $\mu_{d1} - \mu_{d2} > 0$ it follows that

$$\frac{1}{n_{\text{sum}}} \sum_{d \in \boldsymbol{D}} \sum_{i \in \boldsymbol{I}} \frac{{\lambda_{di}}^*}{\mu_{di}} (\mu_{d1} - \mu_{d2}) < \mu_{d1} - \mu_{d2}. \tag{A16}$$

Equation (A12) is established, and Proposition A2 is proven.

Therefore, the objective function of the lower-level problem is equivalent to

$$\min_{\lambda_{di}} Z_{low} = \frac{n_{d1} n_{d2}}{W s_{d1}} - \frac{n_{d1} n_{d2}}{W s_{d2}}. \tag{A17}$$

It is evident that Equation (A17) decreases as λ_{d1} increases. The optimal solution to the lower-level problem is as follows:

$$\begin{cases} {\lambda_{d1}}^{(0)} = (p_1 + p_3)\lambda_d, \\ {\lambda_{d2}}^{(0)} = p_2 \lambda_d. \end{cases} \tag{A18}$$

Substituting Equation (A18) into the upper-level problem, we enter the first iteration. According to Appendix B.1 and Equations (A10) and (A11), it follows that ${n_{di}}^{(1)} = {n_{di}}^{(0)}$. The solution to this bilevel programming problem converges.

The solution to the relaxed bilevel programming problem converges and is given by Equation (A5), subject to Equation (A11).

The results of the relaxed problem of the original model indicate that the upper-level decision maker will make decisions such that all ETC-HVs will choose the ETC lanes. Therefore, even though ETC-HVs have lane selection rights, the ratio of ETC-HVs to CAVs has a homogeneous impact on the relaxed solution. According to Equation (3), for the relaxed solution, only the proportion of MTC-HVs needs to be considered.

Furthermore, in cases where the number of entry and exit lanes at a toll station is specified, such that the total number of lanes in the same direction is fixed, the lane allocation optimization for each direction becomes independent. If we set the arrival rate λ_d of vehicles in one direction to zero, it effectively imposes a constraint that the sum of lanes in the opposite direction is fixed. This allows for the determination of the optimal solution under such conditions. By combining Equations (3) and (A5), we obtain the following result:

$$\begin{cases} n_{d1}^c = \frac{n_d}{1 + \frac{\mu_{d1} p_2}{\mu_{d2}(1 - p_2)}}, \\ n_{d2}^c = \frac{n_d}{1 + \frac{\mu_{d2}(1 - p_2)}{\mu_{d1} p_2}}. \end{cases} \tag{A19}$$

According to Equation (A19), when the number of lanes in the same direction is fixed, the relaxed solution is independent of the vehicle arrival rate. □

Appendix C. Optimal Solutions Under Varying Vehicle Compositions and Arrival Rate

In Figures A1 and A2, the number in each cell represents the optimal solution for lane allocation under the corresponding conditions (determined by the coordinate axis). To improve readability, the color of the cells changes from light to dark as the number of

lanes increases. Cells without numbers indicate that there is no optimal solution under the conditions.

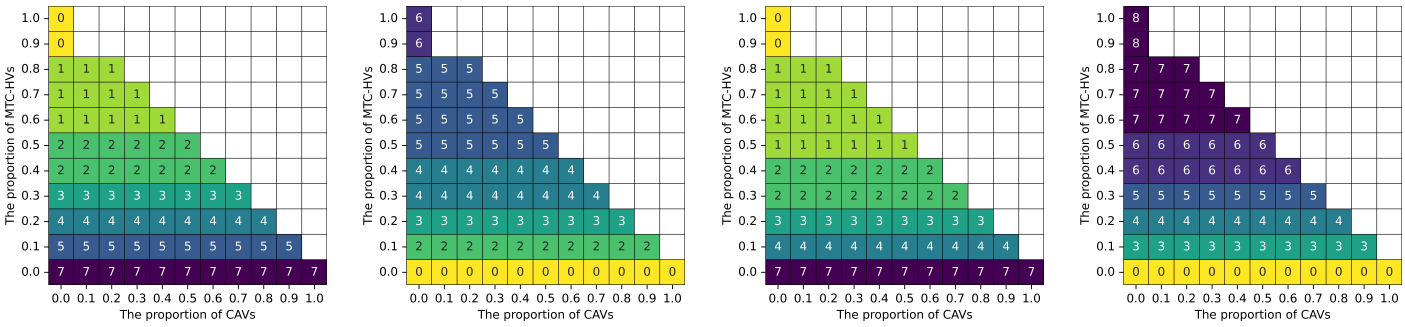

Figure A1. Optimal solutions under varying vehicle compositions. (**a**) Entrance ETC lanes (n_{11}); (**b**) entrance MTC lanes (n_{12}); (**c**) exit ETC lanes (n_{21}); (**d**) exit MTC lanes (n_{22}).

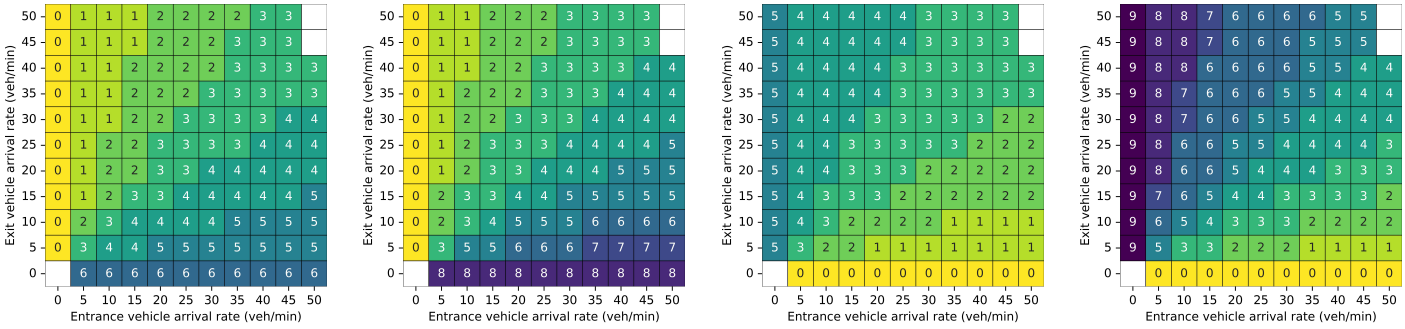

Figure A2. Optimal solutions under varying vehicle arrival rate. (**a**) Entrance ETC lanes (n_{11}); (**b**) entrance MTC lanes (n_{12}); (**c**) exit ETC lanes (n_{21}); (**d**) exit MTC lanes (n_{22}).

References

1. Ahmed, A.H.; Fragonara, L.Z. Adaptive intelligent traffic control systems for improving traffic quality and congestion in smart cities. *Int. J. Qual. Res.* **2021**, *15*, 139–154. [CrossRef]
2. Bansal, P.; Graham, D.J. Congestion in cities: Can road capacity expansions provide a solution? *Transp. Res. Part A Policy Pract.* **2023**, *174*, 103726.
3. Ren, Q.; He, J.; Liu, Z.; Xu, M. Traffic flow characteristics and traffic conflict analysis in the downstream area of expressway toll station based on vehicle trajectory data. *Asian Transp. Stud.* **2024**, *10*, 100138. [CrossRef]
4. Yang, X.; Zou, Y.; Tang, J.; Liang, J.; Ijaz, M. Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models. *J. Adv. Transp.* **2020**, *2020*, 9628957. [CrossRef]
5. Smith, M.; Huang, W.; Viti, F.; Tampère, C.M.; Lo, H.K. Quasi-dynamic traffic assignment with spatial queueing, control and blocking back. *Transp. Res. Part B Methodol.* **2019**, *122*, 140–166. [CrossRef]
6. He, H.-D.; Lu, D.-N.; Zhao, H.-M.; Peng, Z.-R. Characterizing CO_2 and NO_x emission of vehicles crossing toll stations in highway. *Transp. Res. Part D Transp. Environ.* **2024**, *126*, 104024. [CrossRef]
7. Bao, Y.; Xiao, F.; Gao, Z.; Gao, Z. Investigation of the traffic congestion during public holiday and the impact of the toll-exemption policy. *Transp. Res. Part B Methodol.* **2017**, *104*, 58–81. [CrossRef]
8. Liu, C.; Tao, S.; Zhao, C.; Ji, Y.; Du, Y. Optimal deployment of electronic toll collection lanes for freeway network. *China J. Highw. Transp.* **2022**, *35*, 179–188.
9. Anderson, M.L.; Davis, L.W. An empirical test of hypercongestion in highway bottlenecks. *J. Public Econ.* **2020**, *187*, 104197. [CrossRef]
10. Deepashree, R.; Radhika, K.N. Ways to Avoid Traffic Congestion in India and make India Smarter—A Prelude. *Int. J. Eng. Res. Technol.* **2020**, *8*, 209–213.
11. Tseng, P.; Pilcher, N. Political and technical complexities of electronic toll collection: Lessons from Taiwan. *Case Stud. Transp. Policy* **2022**, *10*, 444–453. [CrossRef]

12. Marina, M.; Nikolić, M.; Glavić, D. Optimization of toll road lane operation: Serbian case study. *Oper. Res.* **2022**, *22*, 5297–5322.
13. Petrović, A.; Nikolić, M.; Bugarić, U.; Delibašić, B.; Lio, P. Controlling highway toll stations using deep learning, queuing theory, and differential evolution. *Eng. Appl. Artif. Intell.* **2023**, *119*, 105683. [CrossRef]
14. Zhang, H.; Cheng, C.; Zhang, C.; He, J.; Xu, Y.; Chen, Y.; Hong, Q. Optimization of Opening Scheme of ETC/MTC Toll Lane Based on Cost and Benefit Analysis. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *638*, 012032. [CrossRef]
15. Elassy, M.; Al-Hattab, M.; Takruri, M.; Badawi, S. Intelligent transportation systems for sustainable smart cities. *Transp. Eng.* **2024**, *16*, 100252. [CrossRef]
16. Ngoduy, D.; Nguyen, C.H.P.; Lee, S.; Zheng, Z.; Lo, H.K. A dynamic system optimal dedicated lane design for connected and autonomous vehicles in a heterogeneous urban transport network. *Transp. Res. Part E Logist. Transp. Rev.* **2024**, *186*, 103562. [CrossRef]
17. Di, Y.; Zhang, W.; Ding, H.; Zheng, X.; Ran, B. Cooperative control of dynamic CAV dedicated lanes and vehicle active lane changing in expressway bottleneck areas. *Phys. A Stat. Mech. Its Appl.* **2024**, *638*, 129623. [CrossRef]
18. Almukhalfi, H.; Noor, A.; Noor, T.H. Traffic management approaches using machine learning and deep learning techniques: A survey. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108147. [CrossRef]
19. Saad, M.; Abdel-Aty, M.; Lee, J. Analysis of driving behavior at expressway toll plazas. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *61*, 163–177. [CrossRef]
20. Xiang, W.; Wang, C.; Li, X.; Xue, Q.; Liu, X. Optimizing guidance signage system to improve drivers' lane-changing behavior at the expressway toll plaza. *Transp. Res. Part F Traffic Psychol. Behav.* **2022**, *90*, 382–396. [CrossRef]
21. Subraveti, H.H.S.N.; Srivastava, A.; Ahn, S.; Knoop, V.L.; van Arem, B. On lane assignment of connected automated vehicles: Strategies to improve traffic flow at diverge and weave bottlenecks. *Transp. Res. Part C Emerg. Technol.* **2021**, *127*, 103126. [CrossRef]
22. Wang, P.; Zhao, J.; Gao, Y.; Sotelo, M.A.; Li, Z. Lane Work-Schedule of Toll Station Based on Queuing Theory and PSO-LSTM Model. *IEEE Access* **2020**, *8*, 84434–84443. [CrossRef]
23. Abdelwahab, H.T. Traffic micro-simulation model for design and operational analysis of barrier toll stations. *Ain Shams Eng. J.* **2017**, *8*, 507–513. [CrossRef]
24. Sewagegn, A. Optimization of toll services using queuing theory in the case of ethiopia. *Perner's Contacts* **2022**, *17*, 2. [CrossRef]
25. Qiu, M.; Li, D.; Zhou, D. Determination Optimal Indicators of Expressway Toll Plaza with $M/G/1$ Queue Model. *Int. J. Appl. Phys. Math.* **2020**, *10*, 96–111. [CrossRef]
26. Zhu, D.; Dong, F.; Shi, W.; Zheng, S. The Model of Toll Station Planing. In Proceedings of the 2017 2nd International Conference on Control, Automation and Artificial Intelligence (CAAI2017), Sanya, China, 25–26 June 2017; Volume 134, pp. 390–394.
27. Yuan, N.; Ma, M.; Liang, S.; Wang, W.; Zhang, H. Optimal control method of freeway based on tollbooths lane configuration and variable speed limit control. *Phys. A Stat. Mech. Its Appl.* **2022**, *603*, 127801. [CrossRef]
28. Ramchandra, N.R.; Rajabhushanam, C. Machine learning algorithms performance evaluation in traffic flow prediction. *Mater. Today Proc.* **2022**, *51*, 1046–1050. [CrossRef]
29. Chen, L.-W.; Chen, D.-E. Exploring spatiotemporal mobilities of highway traffic flows for precise travel time estimation and prediction based on electronic toll collection data. *Veh. Commun.* **2021**, *30*, 100356. [CrossRef]
30. *CJJ 37-2012*; Code for Design of Urban Road Engineering. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2016.